doi:10.1038/nature11243

Supplementary Table of Contents

Supplementary Table 1: Number of uniquely mapped monoclonal reads for each ChIP-Seq experiment.

Supplementary Table 2: Pearson correlation coefficient for each pair of biological replicates in ChIP-Seq and RNA-Seq experiments.

Supplementary Table 3: Numbers of expressed genes by BioGPS in each tissue whose promoters are not recovered in this study.

Supplementary Table 4: Promoter usage for RefSeq annotated genes. In this table, we listed the promoter usage for all RefSeq annotated promoters, including the alternative promoters. A promoter is defined as "active" if there is a predicted polII binding site located within 1,000 bp.

Supplementary Table 5: List of 8,792 EPUs defined in this study.

Supplementary Table 6: Table describing the general features of EPUs identified.

Supplementary Table 7: List of linked enhancer promoter pairs in each tissue/cell type.

Supplementary Table 8: List of enhancer promoter pairs tested by 3C and their correlation scores.

Supplementary Table 9: List of the *de novo* motifs found in 19 clusters of tissue-specific enhancer regions. In this table, we listed 206 *de novo* motifs that were found in the tissue-specific enhancer regions. Enhancers were clustered based on H3K4me1 intensity. HOMER *de novo* motif finding software was run on the center of 2kb regions of enhancers with the following parameters: *findMotifsGenome.pl peak_file mm9 output_directory -size 2000 -len 8.* Only motifs with *P* value less than 1e-20 were kept for further analysis. To compare the similarity of *de novo* motifs and known motifs, we used the TOMTOM program from the MEME software suite. Only human and mouse TF motifs were considered as possible matches. Then, based on tissue of origin and gene expression, we manually picked the best match for each *de novo* motif (if available). We also examined the level of conservations for these motifs. We assigned a conservation index for each *de novo* motif based on z-score.

Supplementary Table 10: List of the enriched motifs and their enrichment *P* value as shown in Fig. 4f. (*P* values are log transformed)

Supplementary Table 11: List of the *de novo* motifs that can be matched to a known TF that has been reported to function in the same tissue.

Supplementary Table 12: List of enriched motifs from Homer in tissue-specific promoters and enhancers.

Supplementary Table 13: Primer sequences and chromosome locations of MEF-specifc, mESC-specific enhancers and random genomic regions used for enhancer reporter assay.

Supplementary Table 14: Primer sequences and chromosome locations of novel promoters predicted in MEF, mESCs, and negative genomic regions used for promoter reporter assay.

Supplementary Table 15: List of 3C primers and their location based on mm9.

Supplementary Table 16: The distribution of 373,169,847 uniquely mapped paired-end reads from two Hi-C experiments. The ligation efficiency was calculated based on the number of interactions that are either >20kb for intra-chromosome reads or inter-chromosome reads.

I. Supplementary Methods

Mouse Tissues and Cell Culture

Adult bone marrow, cerebellum, cortex, heart, intestine, kidney, liver, lung, olfactory bulb, spleen, testis, and thymus were dissected from 8-week old male C57Bl/6 mice (Charles River). Placenta was extracted from C57Bl/6 pregnant mice. E14.5 brain, heart, limb and liver, and mouse embryonic fibroblast (MEF) cells were derived from E14.5 C57Bl/6 mouse embryos. MEF cells were genotyped to select male MEF cells used for this study. Placenta was dissected from pregnant C57Bl/6 mice at E14.5. mESC line Bruce4 was maintained on mitomycin C-inactivated MEF feeder layers in DME containing 15% fetal calf serum, leukemia inhibiting factor, penicillin/streptomycin, L-glutamine and non-essential amino acids. mESCs were passaged on 0.2% gelatin twice to deplete feeder cells before harvest for experiments. Tissues were minced to fine pieces in PBS and fixed with 1% formaldehyde at room temperature for 20 minutes.

ChIP-Seq

ChIP-Seq was carried out as previously described [31] with 500 μg chromatin and 5 μg antibody with the following antibodies, H3K4me3 (Millipore 05-745), polII (Covance, MMS-126R),

H3K4me1 (Abcam, ab8895), H3K27ac (Active motif, 39133), CTCF [32], and P300 (Santa Cruz, sc585). ChIP and input library preparation and sequencing procedures were carried out as described previously according to Illumina protocols with minor modifications (Illumina, San Diego, CA).

RNA-Seq

RNA samples from tissues and primary cells were extracted from Trizol® according to protocol (Invitrogen). polyA+RNA was purified with the Dynabeads mRNA purification kit (Invitrogen). The mRNA libraries were prepared for strand-specific sequencing as described previously [3333].

Promoter and Enhancer reporter assay

Predicted promoter and enhancer sequences were randomly selected for validation in reporter assays. The chromosome coordinates and primers were listed in Supplementary Table 13 and 14. Cloning and reporter assays were carried out as previously described [34]. For novel promoter sequences, we tested both orientations of the candidate sequences. Fragments were designated as active if their relative luciferase value was significantly higher than random genomic fragments (P value < 0.01).

ChIP-Seq and RNA-Seq data processing

ChIP-Seq reads were aligned to the mouse genome build mm9 with Bowtie version 0.12. Some of the early ChIP-Seq results were aligned by ELAND. We used the first 25 bp for the alignment and only kept the reads with less than two mismatches. To generate the wig files, we extended

the mapped reads to 300 bp toward the 3' end, divided the mouse genome into 100 bp bins, and counted the number of reads that fell within each bin. We normalized the tag counts in each bin according to the total number of reads. Input reads were processed in the same way and their normalized signal intensity values were subtracted from the ChIP-Seq tracks. Therefore, the height of each 100 bp bin in genome browser is computed as: Δnormalized signal intensity = normalized signal intensity$_{IP}$ - normalized signal intensity$_{input}$.

For RNA-Seq data, we mapped raw reads in FASTQ format to the mouse genome with TopHat software version 1.20 [35]. The wig files for RNA-Seq data were generated by TopHat. We assigned expression value for each gene in RefSeq with Cufflinks software [36]. To normalize the gene expression levels between different tissues, we used the quantile normalization function in R.

Data reproducibility

To examine the reproducibility of our ChIP-Seq experiments, we performed the following analysis. First, we divided the mouse genome into 1000 bp bins, and computed normalized signal intensity values as described above. Each replicate could be represented as a vector of 2.65 million numbers. We computed the Pearson correlation coefficient between the two biological replicates for every mark in each tissue/cell type and the results were listed in Supplementary Table 2. After we validated that two replicates were highly correlated, we pooled them together for further analysis (Supplementary Fig.12a and b).

Identification of *cis*-regulatory elements

To map promoters, we relied exclusively on the presence of H3K4me3 [34, 37]. To identify enhancers, we took advantage of the chromatin signature pattern that they share, i.e. the presence of H3K4me1 but absence of H3K4me3 [34, 38], and developed a chromatin-signature based

enhancer predictor trained with the distal p300 binding sites in mESCs (Supplementary Fig. 13a). Recent studies show that H3K4me1 marks both poised and active enhancers, while H3K27ac marks active enhancers [39, 40]. Consistent with this finding, we found H3K27ac at only a portion (between 15 and 40%) of enhancers identified in this work (Supplementary Fig. 13c). To identify potential insulator elements, we determined the binding sites of CTCF in each tissue [32]. To accurately analyze the Chip-Seq data, we developed a computational pipeline (Supplementary Fig. 12c). We first identified the potential binding sites with MACS [41] with the default parameter (*P* value < 1e-5). To ensure we had good quality peaks for further analysis and to address the difference in the sequencing depths between different data sets, we performed the following procedures to further filter the peaks. We computed the normalized signal intensity values for the 1kb region centered at the "summit" of peaks predicted by MACS in the ChIP-Seq and input data. Then we applied the following parameters to finalize the enrichment regions for H3K4me3 and CTCF: two-fold enrichment (normalized signal intensity$_{IP}$ >= 2*normalized signal intensity$_{input}$) and Δnormalized signal intensity >1. For testis, we only kept H3K4me3 peaks that overlap with UCSC known genes TSS [42] due to its demonstrated abundance at recombinant hotspot in testis [43]. To predict polII occupancy, we required the peaks to be called by MACS first and also have a Δnormalized signal intensity >1. To predict enhancers, we adopted a previously published method based on the chromatin signatures of H3K4me1 and H3K4me3 [34]. Specifically, we first binned the ChIP-Seq data and input data into 100 bp bin and computed a normalized intensity value for each bin. We collected the H3K4me1 and H3K4me3 peaks around the distal p300 binding sites as the training data set. We used a sliding window to scan the genome comparing the H3K4me1 and H3K4me3 profiles with the training data. We used a discriminative filter on H3K4me1 and H3K4Me3 to keep only those sites that correlated with the averaged enhancer training set more than the promoter training set. Finally, we applied a descriptive filter on

H3K4me1 and H3K4me3, keeping only those remaining predictions having a normalized intensity of at least 0.5.

We defined novel promoters as the H3K4me3 peaks that are at least 3kb away from known gene bodies and compared them with other datasets. 75% of them demonstrated evidence of transcriptional initiation, such as binding to unphosphorylated RNA pol II, or capable of making 5'-capped RNA as suggested by cap analysis of gene expression (CAGE) data [44, 45], or both (Supplementary Fig. 3c).

Conserved usage of *cis*-regulatory elements in the mouse and human genomes

To examine the sequence conservation of the identified *cis*-regulatory elements, we evaluated their PhastCon scores [46]. We randomly chose 1000 exons as positive controls and 1000 random intergenic regions as negative controls. For *cis*-regulatory elements, the highest PhastCon scores in the 500 bp around the center of all elements, except for exons and promoters. For exons, we used the highest score within the exon, and for promoters, we used only 500 bps upstream of TSS. We converted the predicted promoters, enhancers and CTCF binding sites from mouse genomic locations (mm9) to human genomic locations (hg18), using the liftOver tool [47] from UCSC genome browser [48] with the center 200 bp of each element and required minMatch > 0.5. We considered the usage of a *cis*-regulatory element conserved if the corresponding human homologous sequence is bound by the same factor within 2kb regions.

Identification of tissue-specific *cis*-regulatory elements

To quantitatively measure the relative occupancy for each *cis*-regulatory element, we adopted a strategy based Shannon Entropy to assign a tissue-specificity index to each element [49]. Specifically, for *cis*-regulatory elements, we defined its relative occupancy in a tissue t as $p_{t,s} = B_{t,s} / \Sigma_{1 \le t \le N} B_{t,s,}$ where $B_{t,s}$ is calculated as the normalized binding intensity relative to input and N is the number of tissues. The entropy score is defined as $H_s = -1*\Sigma_{1 \le t \le N} p_{t,s} * \log_2(p_{t,s})$, where the value of H_s ranges between 0 to $\log_2(N)$. An entropy score close to zero indicates the occupancy at this site is highly tissue-specific, while an entropy score close to $\log_2(N)$ means the site is bound uniformly. In Fig. 3a, we plotted the tissue-specficity index for each category of the *cis*-regulatory elments, measured by the entorpy score. For the x axis (tissue-specificity), we plotted 2 to the power of entropy score to gain a more intuitive view of the number of tissues. In Supplementary Fig. 14, we used entropy score less than 2.1 to define tissue-specfic polII binding and entropy score greater than 4.0 to define ubiquitous polII occupancy.

For Supplementary Fig5a, we first ranked all the CTCF binding sites according to the tissue-specificity index and defined the top 25% of the list as tissue-specific and the bottom 25% as ubiquitous.

To investigate the relationship between tissue-specific polII binding and gene expression, we plotted the signal intensities of polII binding and gene expression for a subset of RefSeq promoters that show tissue-specific usage. We confirmed that tissue-specific polII binding correlates with tissue-specific gene expression (Supplementary Fig. 14a). Gene ontology analysis shows that genes with tissue-specific polII binding are mostly associated with tissue-specific function (Supplementary Fig. 14b).

Previous studies showed that housekeeping gene promoters are usually associated with high CpG content, while tissue-specific gene promoters tend to associate with low CpG content [50]. To confirm whether it is true in the murine tissue/cell types we studied, we classified 23,523 RefSeq gene promoters into high CpG promoters (HCP), low CpG promoters (LCP) and intermediate CpG promoter (ICP) as previously described [50]. As expected, promoters ubiquitously occupied by polII in 19 tissue/cell types are highly enriched for HCPs, while promoters bound by tissue-specific polII are more enriched for ICPs and LCPs (Supplementary Fig. 14c).

Correlation of enhancers with promoters

First, we compared the predicted enhancers with distal p300 binding sites, and found most of them were recovered by predicted enhancers within 1.5 kb (Supplementary Fig. 15). Therefore, we merged enhancers from different tissue/cell types that are located within 1.5 kb and used the midpoint as the center of the "new" enhancer. For promoters, we computed the Δnormalized polII and H3K27ac intensity of the 1kb window in 19 tissue/cell types, which can be represented by a vector of 19 numbers. Next we computed the vectors of Δnormalized H3K4me1 and H3K27ac signal intensity for each enhancer using a 3kb window. Then we computed the Spearman correlation coefficient (SCC) between an enhancer and a promoter (polII at promoter vs H3K4me1 at enhancer, H3K27ac at promoter vs H3K27ac at enhancer). For the random model in Fig. 3b, we randomly shuffled the Δnormalized signal intensity in each tissue for each enhancer and promoter and randomly assigned a target promoter for each enhancer. The simulation was performed 100 times and results were combined to generate the figure. For the nearest TSS model, we assigned each enhancer to its nearest promoter. For the

CTCF block model, we first divided the mouse genome into blocks based on CTCF binding sites that do not overlap with promoters and enhancers. Enhancers and promoters that are located within the same block were assigned as linked pairs.

Identification of Enhancer-Promoter Unit

Starting from the first element in each chromosome, we calculated its SCC with the next element. If the "new" element was highly correlated (SCC > 0.23) with the current element or with at least 50% of all the elements in the current block, it was added into the current block. Otherwise, we closed the expansion of the current block and started a new block with that new element. We only kept EPUs with at least one promoter and one enhancer. All the possible promoter and enhancer pairs within the same EPU with SCC > 0.23 were defined as linked enhancers and promotes.

Hi-C and 3C experiments

Cortex Hi-C experiments were conducted in biological replicates with HindIII restriction enzyme according to previous publication [51] with modifications for tissue samples. In brief, cortex from 8-week old male C57Bl/6 mice were dissected, minced and fixed with 1% formaldehyde for 20 minutes at room temperature. After fixation, samples were homogenized and counted for cell numbers with Trypan Blue. We obtained about 10-15 million cortex cells from each animal routinely, and 20-30 million cells were used for each experiment.

3C experiments in cortex and mESCs were performed following standard procedures [52, 53] with a few modifications. ~25 million cells from cortex and mESCs were crosslinked as described for Hi-C procedure. Crosslinked cells were lysed and digested with 400 units of HindIII (NEB) overnight at 37^{o}C. The digested chromatin were subsequently ligated with 50 units of T4 DNA

ligase (Invitrogen) at 16 °C. The ligated samples were reserved crosslinked and purified for 3C analysis. Meanwhile 20ug of BAC clones (RP23-69N8 and RP24-369D18 for locus containing Fam13 , RP24-68D20 and RP23-225C22 for locus containing Gucy1a3, RP24-248L13 for locus containing Trim19) covering each region were digested with 400 units of HindIII (NEB) and randomly ligated with 20 units T4 DNA ligase (Invitrgen) at 16 °C overnight to create all possible ligation products. 3C-qPCRs were done in triplicate and the relative interaction frequency for each point was first corrected by PCR efficiency of each primer pair. To compare the differences in interaction frequency between cortex and mESCs, we used the control region in Ercc3 gene [54]. For 3C primers, please see Supplementary Table 15.

Cortex Hi-C data processing

The paired-end Hi-C reads were mapped to the mouse genome build mm9 using an in-house pipeline based on BWA [55]. Duplicated reads from the same biological library were removed. We obtained a total of ~373 million monoclonal paired-end reads from two biological replicates (Supplementary Table 16), of which nearly 60% represent long-range interactions (with both ends at least 20kb away from each other as described previously [51]). The heatmaps for Hi-C interaction frequency were generated as previously described [51]. In specific, we binned the mouse genome into 20kb bins (for Supplementary Fig. 16a, we binned the genome into 200kb bins for displaying purpose) and the Hi-C interaction frequency $I_{i,j}$ between bin i and bin j is defined as the number of paired-end reads that mapped from bin i to bin j. We found that Hi-C experiments are highly reproducible (Supplementary Fig. 16a) and the interaction frequency matrices are strongly correlated (Pearson correlation coefficient = 0.98, Supplementary Fig. 16b). The data from two experiments were pooled together for further analysis. We also performed

extensive data quality control and data normalization as described previously [56] and in Dixon et al., 2012 [57]. For Fig. 3e, we only used enhancers and promoters that are active in cortex and required them to be located in different bins.

Clustering analysis

We performed clustering analysis on the entire set of enhancers identified in this study with Cluster 3.0 software (Fig. 4d) [58]. The color scheme for each cell is based on the Δnormalized signal intensity of a 3kb region around the enhancer. Promoter activities were measured by the Δnormalized signal intensity of a 1kb region around the TSS and gene expression levels were calculated as the RPKM values from the RNA-Seq experiments. For Fig. 4f we performed hierarchical clustering with tissues arranged by Kendall's tau and motifs grouped by correlation.

Motif analysis

We first identified 19 clusters of tissue-specific enhancers (Fig. 4e), and ran HOMER [59] *de novo* motif finding software on the center 2kb regions with the following parameters: *findMotifsGenome.pl peak_file mm9 output_directory -size 2000 -len 8*. Only motifs with *P* value < 1e-20 were kept for further analysis. To identify the *de novo* motifs with known transcription factor motifs in the mammalian genome, we used the TOMTOM program [60] from the MEME software suite (http://meme.sdsc.edu/meme4_3_0/cgi-bin/tomtom.cgi). We chose TRANSFAC, JASPAR CORE and UNIPROBE as the candidate databases. The comparison function was set as "Pearson correlation coefficient". Typically, there were multiple candidates for each *de novo* motif. In an effort to identify the most likely transcription factor, we filtered the candidates by choosing a cutoff *P* value < 0.0005 and an empirical FDR < 0.02. To compute the

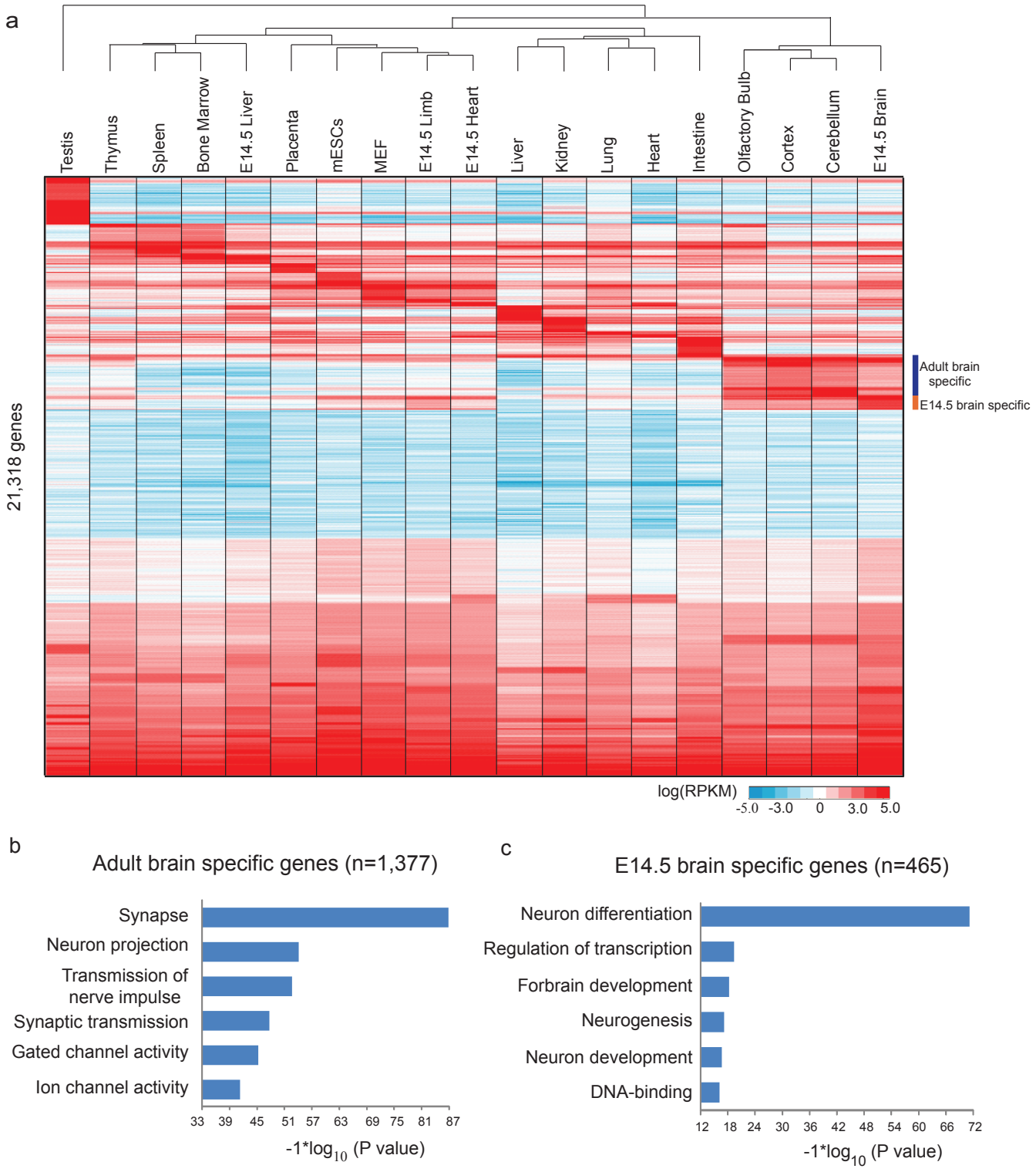
FDR for each motif, we shuffled the motif 1000 times and run TOMTOM with the pseudo motifs against the known motif database. The FDR was computed as the rank of the original *P* value among the *P* values for the randomly generated motifs. Next we manually inspected the remaining candidates, requiring that the candidate transcription factor is expressed in the same tissue. To perform the motif enrichment analysis, we combined all *de novo* motifs and the known motifs from HOMER [59]. Only the motifs with an enrichment *P* value < 1e-20 in at least one cluster of enhancers were presented in Fig. 4f.

To perform the motif conservation study, we first located the motif occurrences in the tissue-specific enhancers. The conservation score was computed as the sum of the average of phastCon scores at each base pair for all motif occurrences. To compare, we randomly generated 1000 8-mers and compute their average PhastCon score. We repeated this step 1000 times to generate a population of the phasCcon scores. Then based on the average and standard deviation of this population, we computed a Z-score for each de novo motif. We defined a motif as conserved when z-score > 2.58.

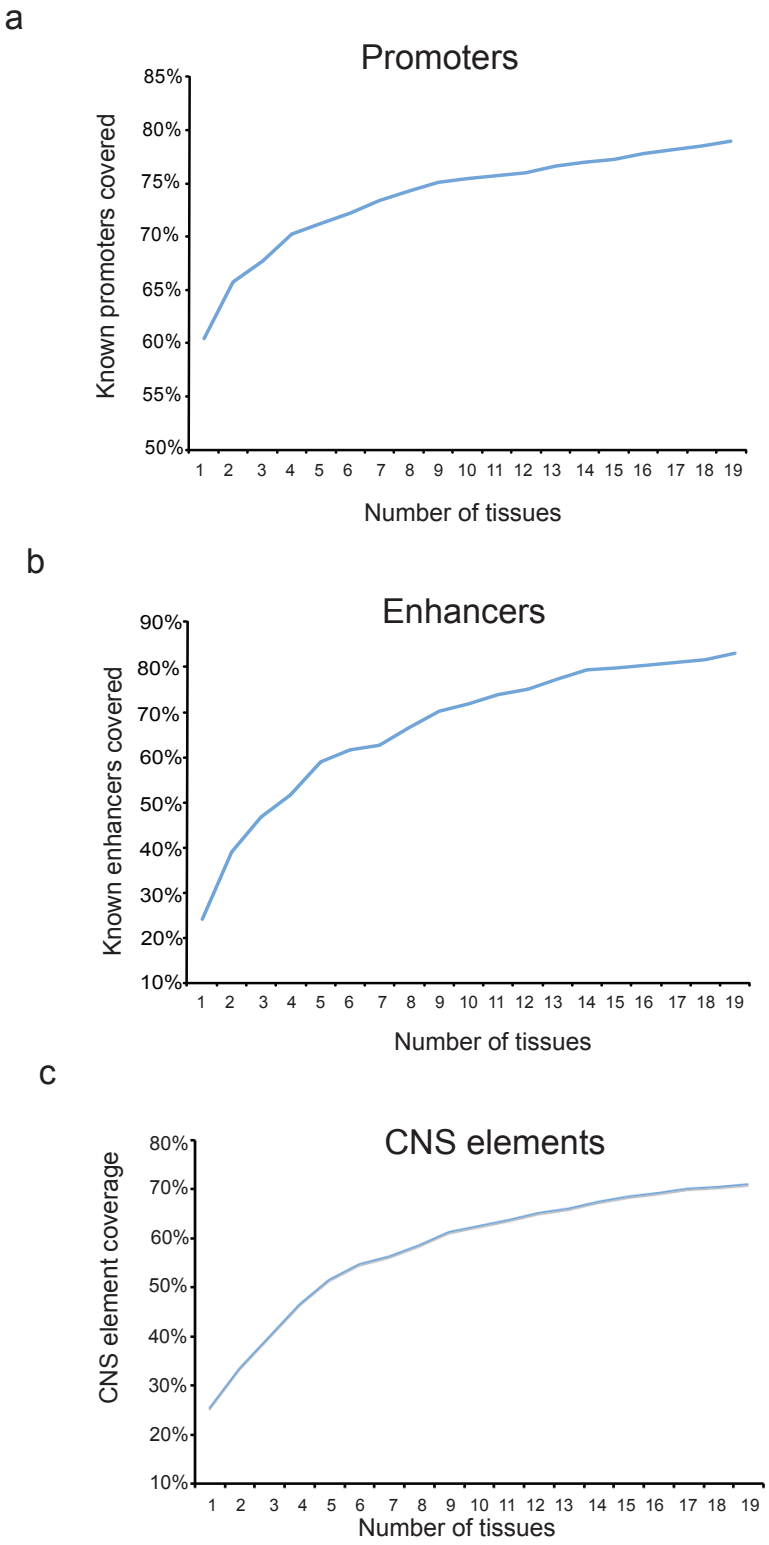
II. Supplementary References

31. Hawkins, R.D. et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479-91.
32. Kim, T.H. et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-45 (2007).
33. Parkhomchuk, D. et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37**, e123 (2009).
34. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
35. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
36. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5.
37. Bernstein, B.E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-81 (2005).
38. Heintzman, N.D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-12 (2009).

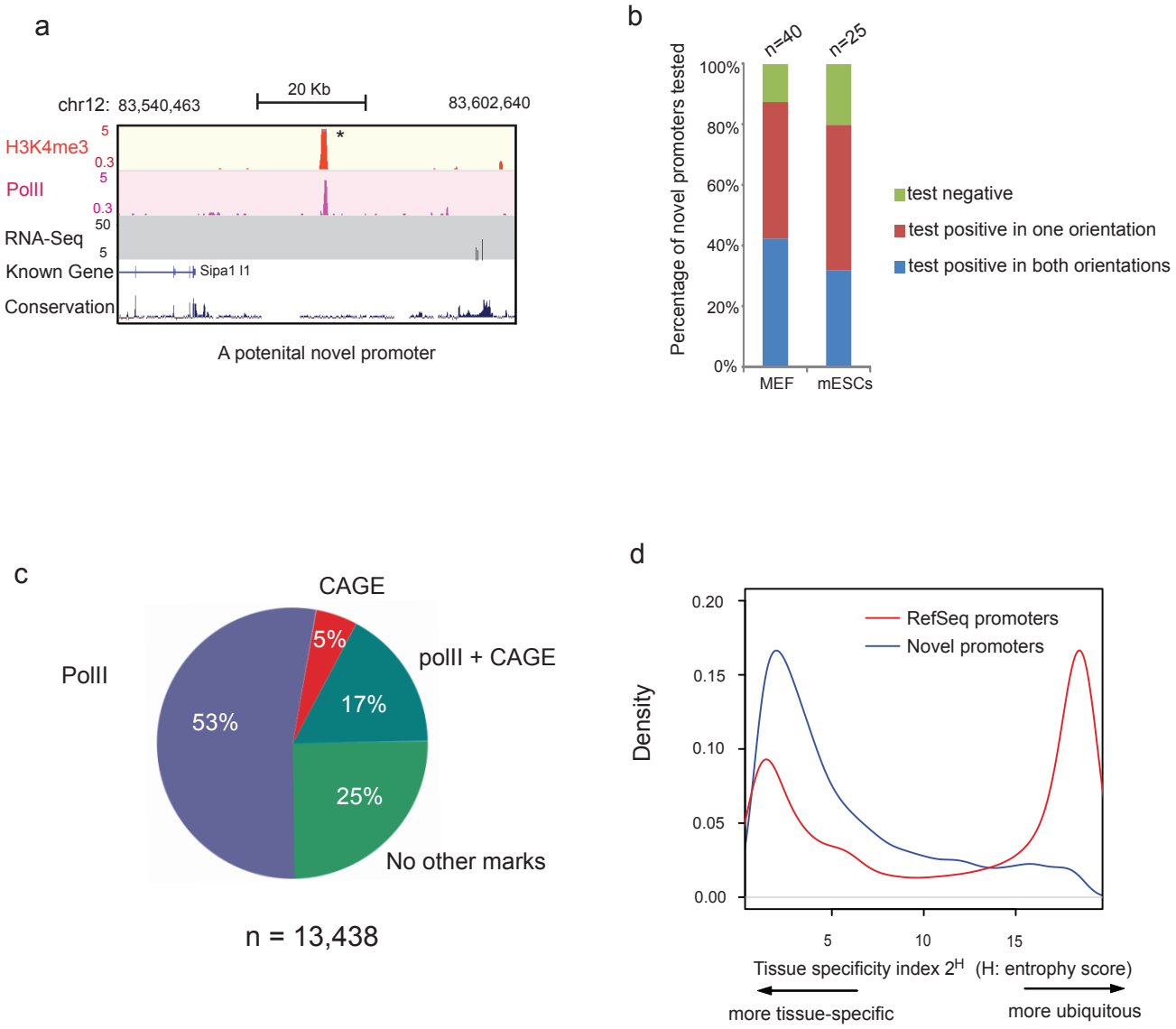
39. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6.
40. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83.
41. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
42. Hsu, F. et al. The UCSC Known Genes. *Bioinformatics* **22**, 1036-46 (2006).
43. Smagulova, F. et al. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* **472**, 375-8.
44. Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-35 (2006).
45. Shimokawa, K. et al. Large-scale clustering of CAGE tag expression data. *BMC Bioinformatics* **8**, 161 (2007).
46. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-50 (2005).
47. Hinrichs, A.S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-8 (2006).
48. Fujita, P.A. et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39**, D876-82.
49. Barrera, L.O. et al. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* **18**, 46-59 (2008).
50. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
51. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
52. Miele, A. & Dekker, J. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods Mol Biol* **464**, 105-21 (2009).
53. Hagege, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722-33 (2007).
54. Splinter, E. et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**, 2349-54 (2006).
55. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
56. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* **43**, 1059-65.
57. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*.
58. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-4 (2004).
59. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89.
60. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007).



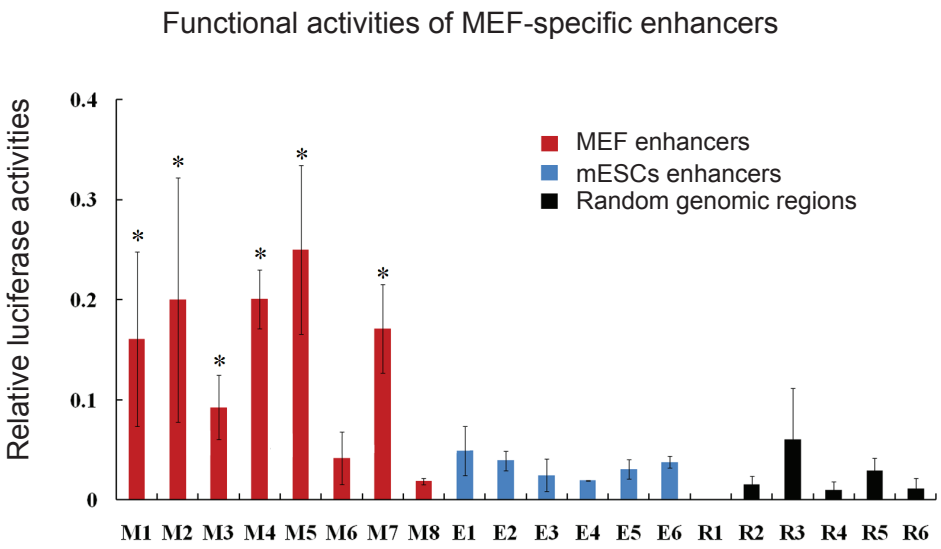
Supplementary Figure 1: Gene expression profiles in nineteen mouse tissues and cell types measured by RNA-seq. a, A heatmap showing 21,318 RefSeq genes that were expressed in at least one tissue (RPKM >1). **b** and **c**, GO analysis for adult brain specific genes (highlighted by blue bar) and embryonic brain specific genes (highlighted by orange bar).



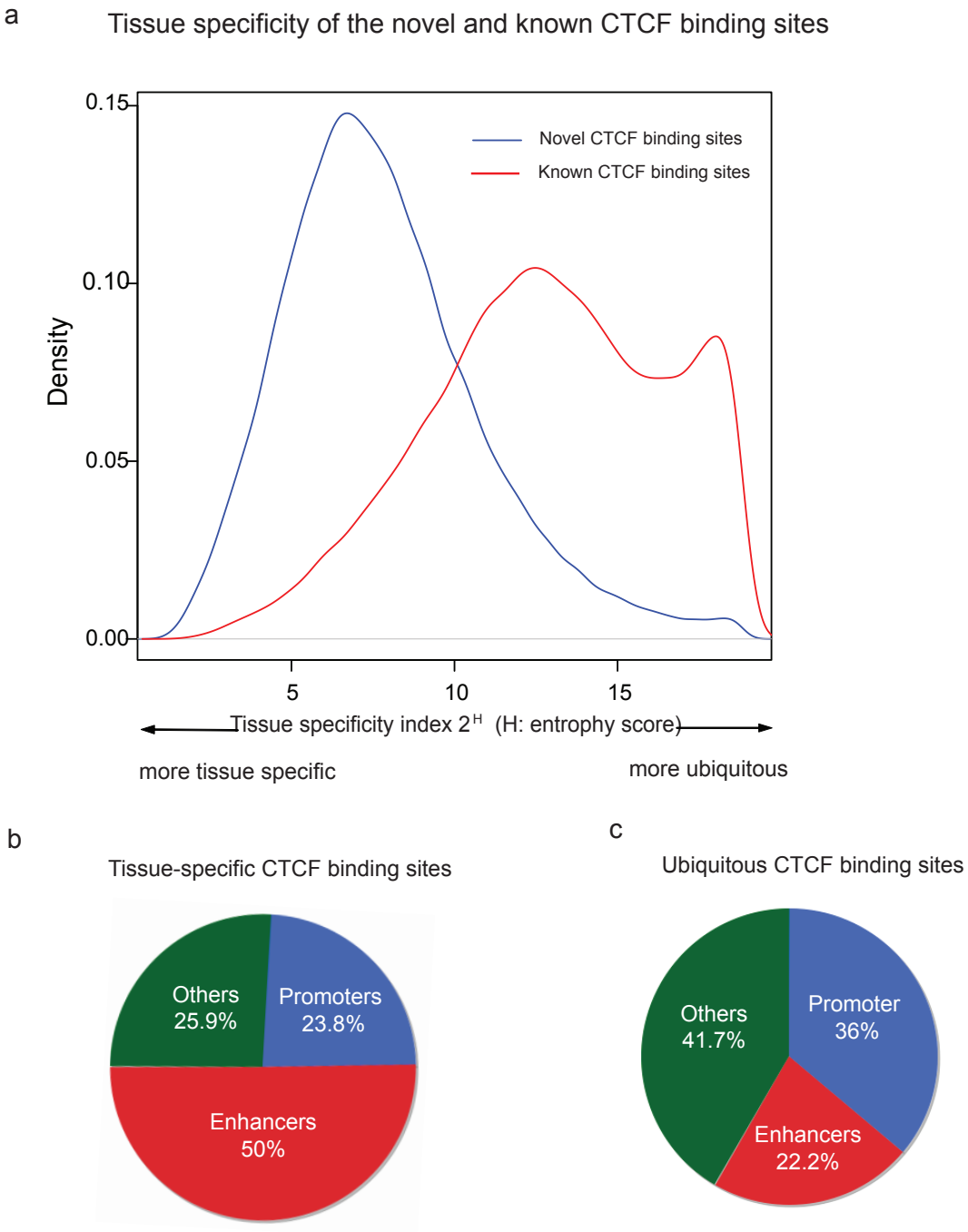
Supplementary Figure 2: Saturation analysis of promoters, enhancers and conserved non-coding sequence (CNS) elements. a, Percentage of RefSeq annotated promoters recovered by our method, by using increasing number of tissue and primary cell types. **b**, Percentage of known enhancers recovered by our method, by using increasing number of tissue and primary cell types. **c**, Percentage of CNS elements recovered by the cis-regulatory elements identified in this study, by using increasing number of tissue and cell types.



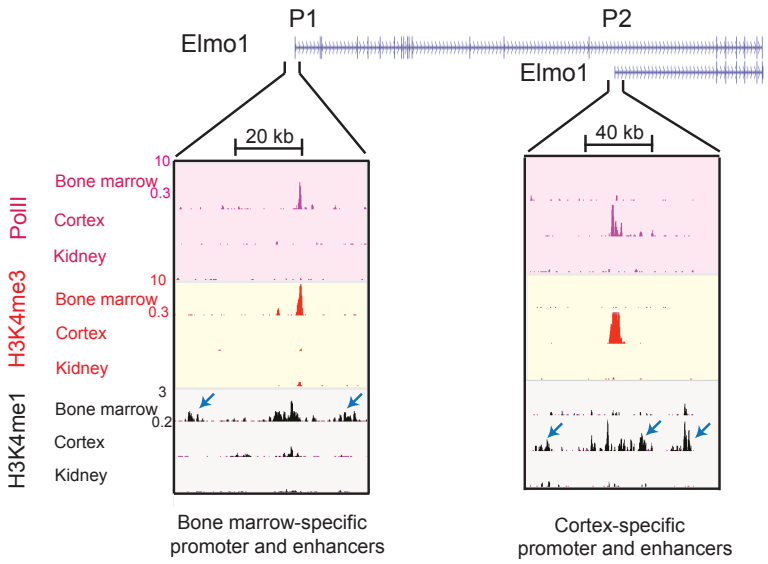
Supplementary Figure 3: Novel promoters identified in this study. a, An example of a predicted novel promoter in mESCs. **b**, Luciferase reporter assay results of novel promoters identified in MEF and mESCs. **c**, Most novel promoters are supported by other datasets, including CAGE and polII binding. **d**, Novel promoters are more tissue-specific than RefSeq annotated promoters.



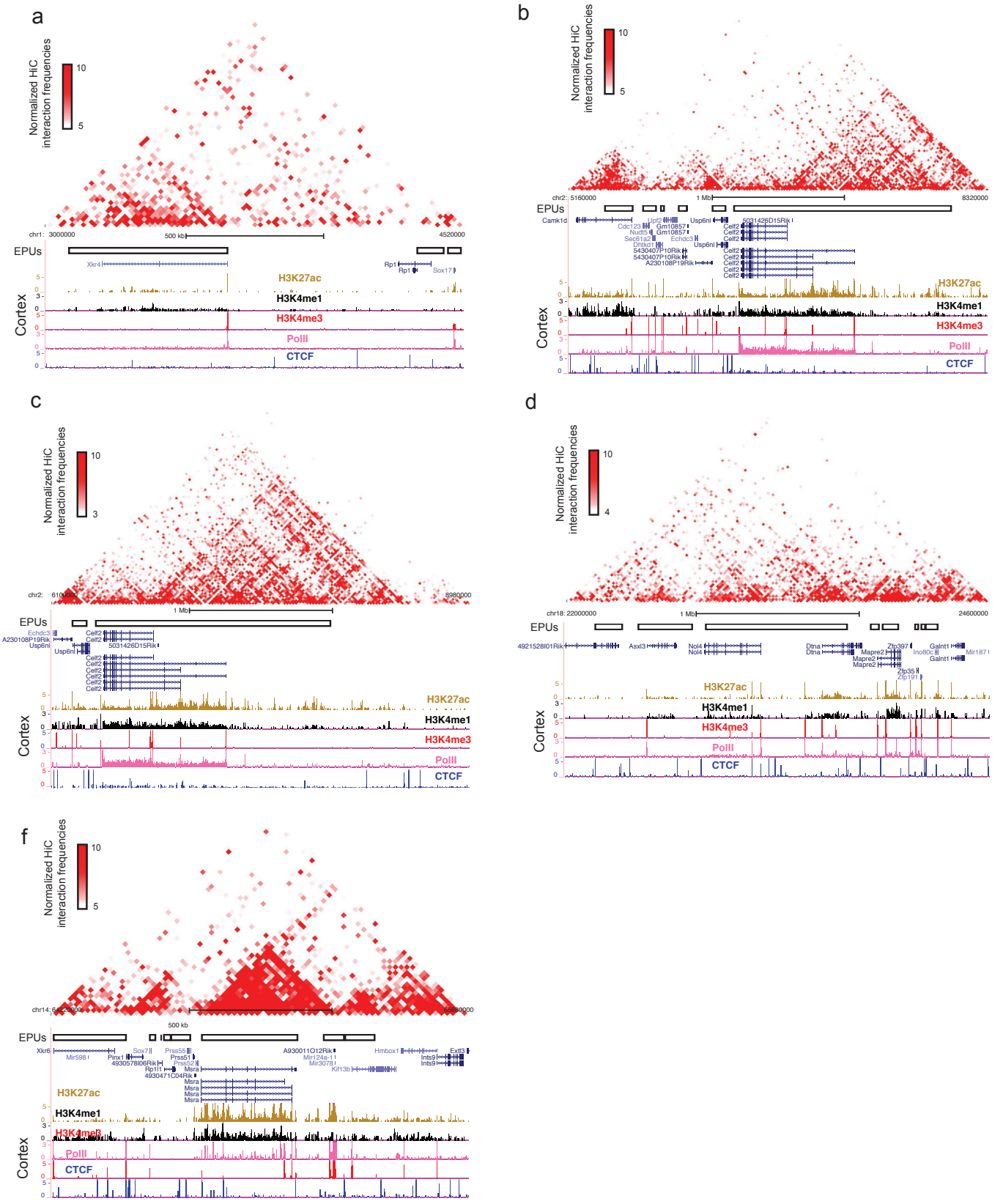
Supplementary Figure 4: Chart showing luciferase reporter assay results of eight MEF-specific enhancers. Luciferase assays in MEF show that MEF-specific enhancer sequences (M1-M8) drive the reporter expression significantly better than a set of six random genomic regions (R1-R6) and six mESC specific enhancers (E1-E6). Error bars indicate the standard deviation of three independent report assays. (* P value < 0.01, T test).



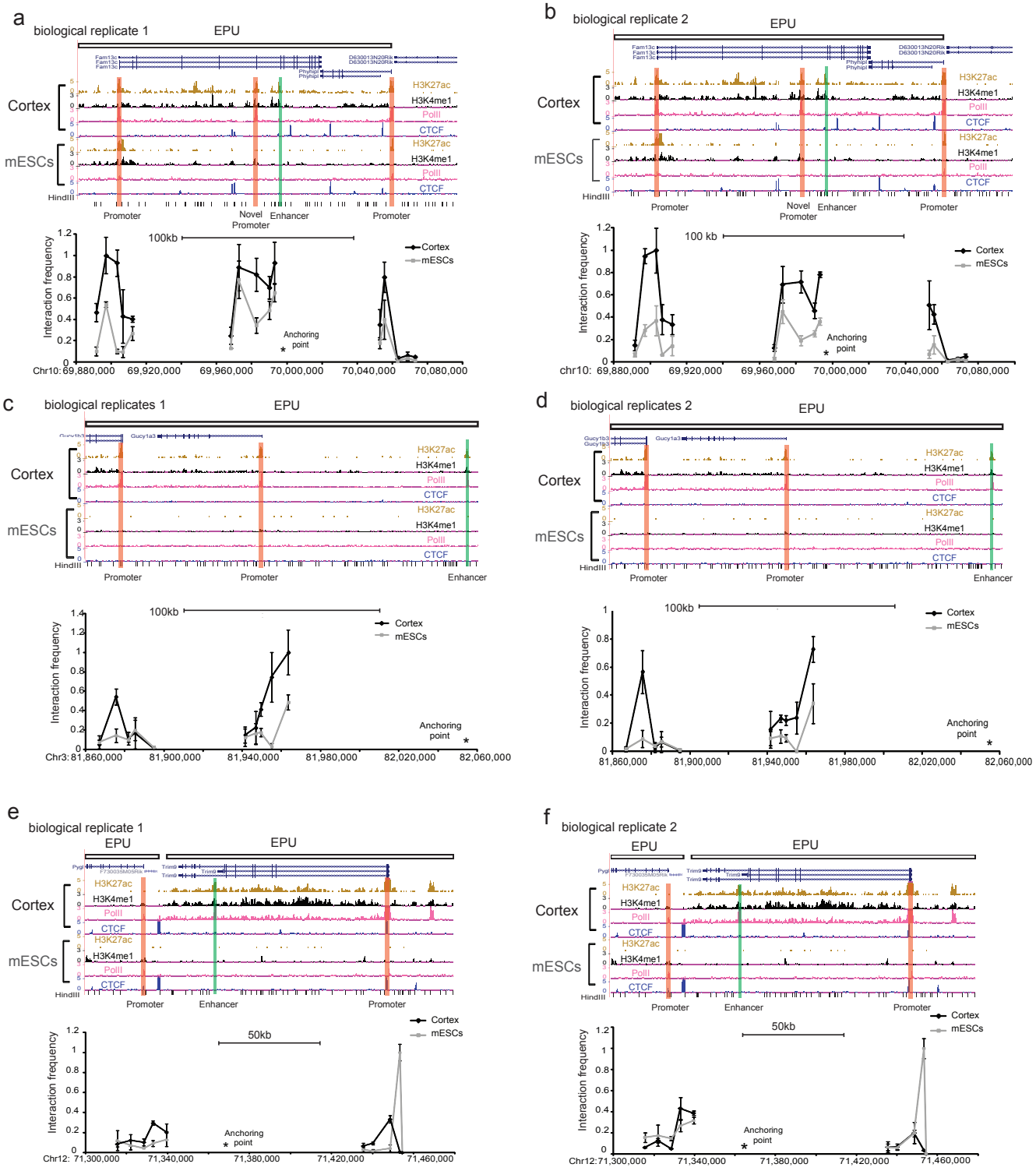
Supplementary Figure 5: Tissue-specificity of CTCF binding sites and the comparison of CTCF binding sites with promoters and enhancers. a, Tissue specificity of the novel CTCF binding sites identified in this study. Novel CTCF binding sites are more tissue-specific than known CTCF binding sites (P-value < 2.2e-16, Wilcoxon test). **b**, Percentage of tissue-specific CTCF binding sites that overlap with promoters and enhancers. **c**, Percentage of ubiquitous CTCF binding sites that overlap with promoters and enhancers.



Supplementary Figure 6: Tissue-specific occupancies of polII at alternative promoters are correlated with H3K4me1 chromatin marks at enhancers. We observed that the chromatin states of enhancers are correlated with polII occupancies at two alternative promoters (P1 and P2) for the Elmo1 gene. As indicted by the polII signals, P1 is active in cortex, while P2 is active in bone marrow. Interestingly, we also observed bone marrow and cortex specific enhancers in this region.

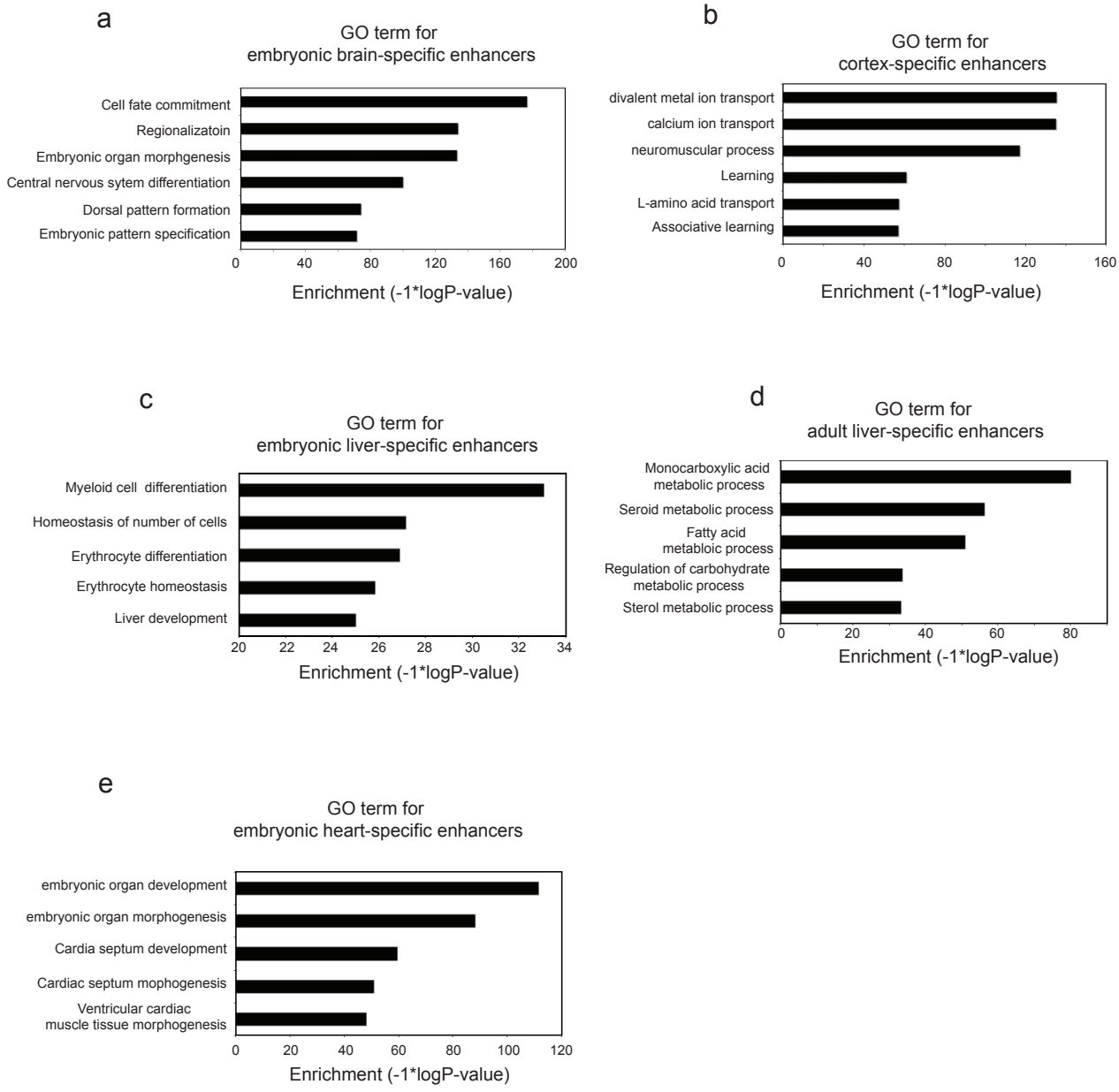


Supplementary Figure 7: Comparing EPU blocks and Hi-C interaction frequencies in cortex. a-e, (Top) Normalized Hi-C interaction frequencies in mouse cortex as a two-dimensional heatmap. (Bottom) UCSC genome browser views of the same regions, including the identified EPUs and the ChIP-Seq data (H3K27ac, H3K4me1, H3K4me3, polII and CTCF).

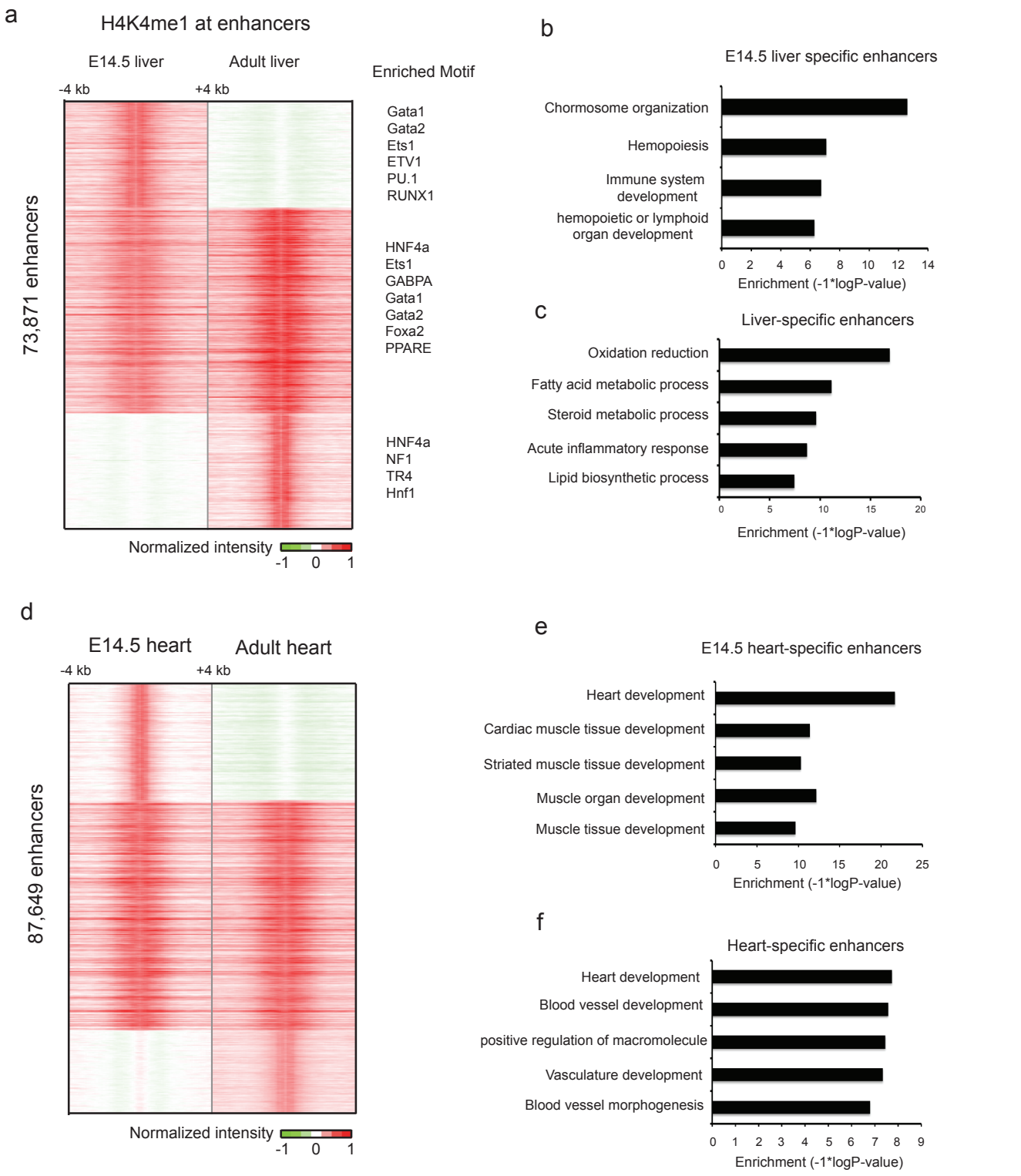


Supplementary Figure 8: Validation of linked promoter and enhancer pairs in cortex by 3C. DNA looping events were detected between enhancers (anchoring points) and promoters in cortex, but not in mESCs. HindIII cutting sites are shown above the 3C graphs. As shown in **a** and **b**, an enhancer interacts with the upstream *Fam13c* promoter located within the same EPU, but not with the *Phyhipl* gene promoter which is at a closer distance but outside the EPU. We

also noticed that this enhancer interacts with a putative novel promoter in the same EPU. The correlation between the enhancer and *Phyhipl* gene promoter is low (Spearman correlation coefficients of 0.08 between the H3K4me1 intensity at enhancer and polII at promoter, and 0.04 between H3K27ac at enhancer and H3K27ac at promoter). The correlations of the two linked promoter/enhancer pairs are much higher (0.4 and 0.25 for H3K4me1/polII respectively, and 0.5 and 0.34 for H3K27ac/H3K27ac respectively. In **c** and **d**, an enhancer interacts with a distal gene *Gucy1b3* promoter within the same EPU (Spearman correlation coefficients of 0.45 for H3K4me1/polII and 0.6 for H3K27ac/H3K27ac), "bypassing" the promoter of a neighboring gene *Gucy1a3* (Spearman correlation coefficients of 0.45 for H3K4me1/polII and 0.6 for H3K27ac/H3K27ac). **e** and **f**, an enhancer (anchoring point) interacts with the *Trim9* gene promoter which is located within the same EPU, but not with the *Pyg1* gene promoter which is located at closer genomic distance but outside the EPU. The detected peak downstream of *Trim9* gene in mESCs is due to unknown mechanism.



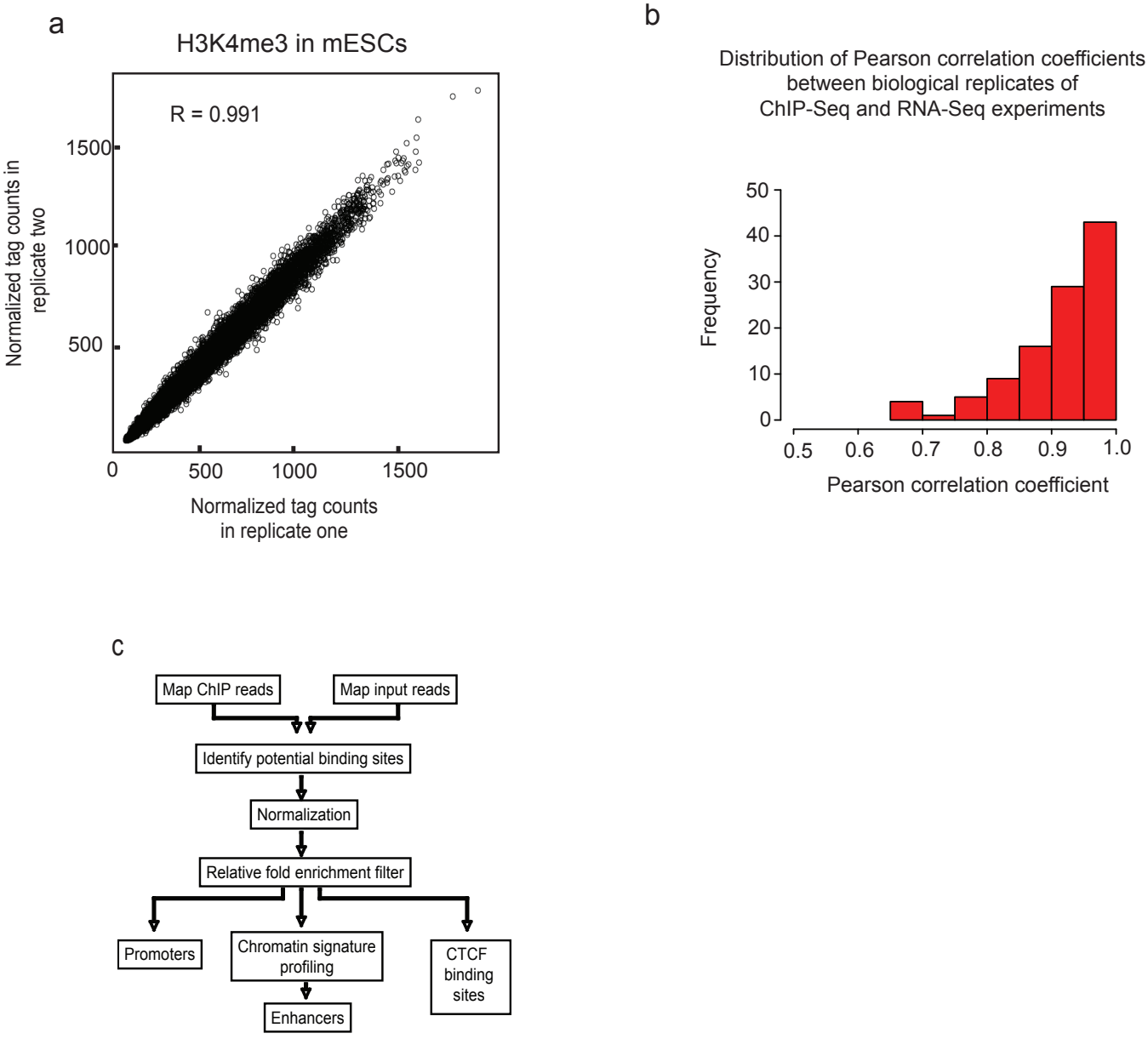
Supplementary Figure 9: Gene Ontology analysis for embryonic and adult stage-specific enhancers by GREAT. a-e, We used default settings for GREAT. The domain model was set as "basal plus extension" and the proximal was defined as 5kb upstream and 1kb downstream, plus Distal up to 1 Mb. Test regions are the +/- 100 bp around the center of the predicted enhancers and the background regions are the whole genome. Redundant GO categories were removed.



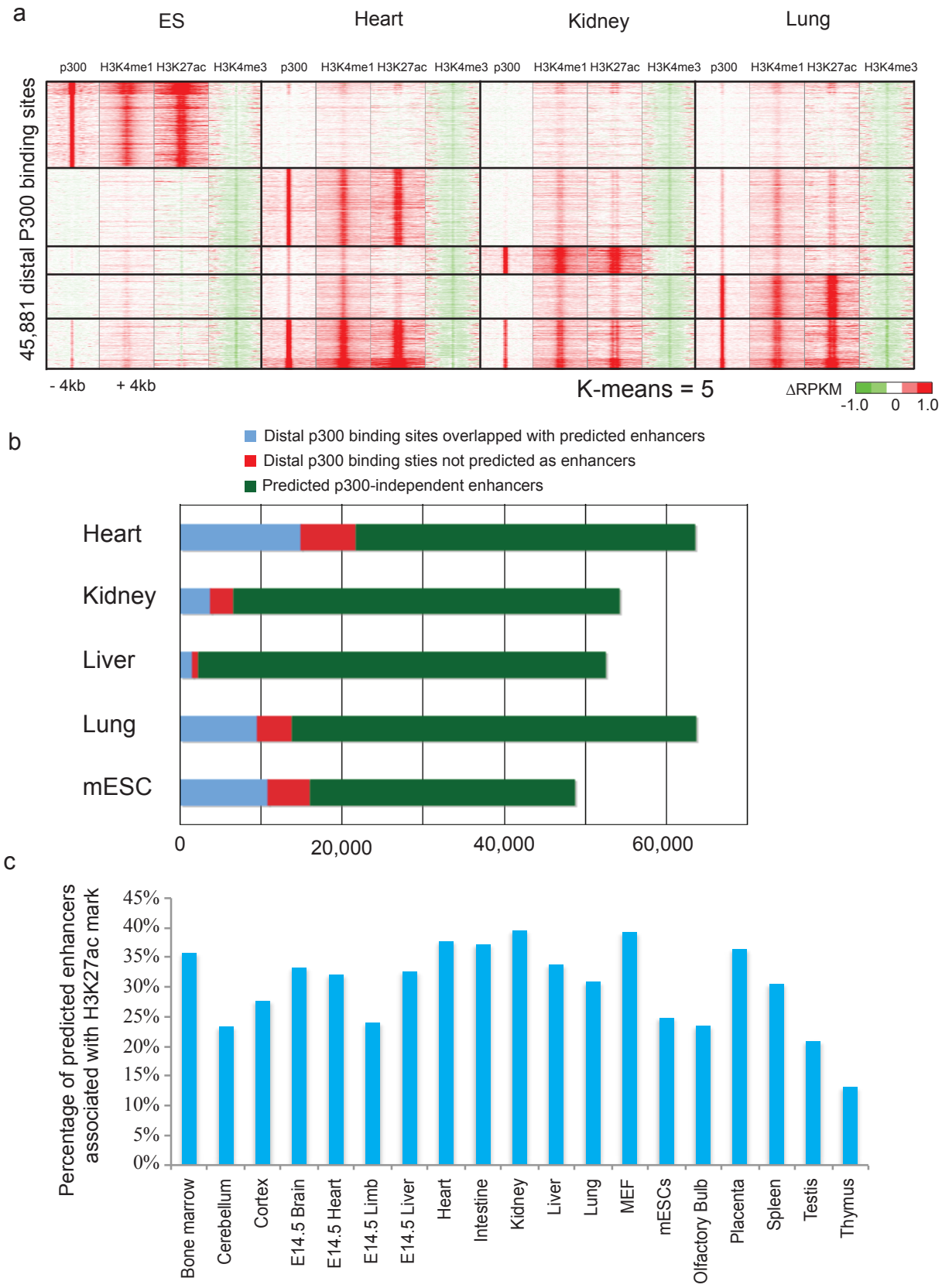
Supplementary Figure 10: Developmental stage-specific enhancers in liver and heart. **a**, Classification of developmental stage-specific enhancers based on their chromatin state (H3K4me1). In addition, motifs enriched in development stage-specific liver enhancers are annotated on the right side. **b** and **c**, Gene Ontology analysis for the genes associated with embryonic liver-specific and adult liver-specific enhancers by EPU analysis. **d**, Classification of developmental stage-specific enhancers for embryonic and adult heart. **e** and **f**, Gene Ontology analysis for the genes associated with embryonic heart and adult heart specific enhancers by EPU analysis.


Bone marrow
regulation of cytokine production
regulation of cell activation
leukocyte chemotaxis
positive regulation of cytokine biosynthetic process
positive regulation of myeloid leukocte differentiation
0 5 10 15 20 25
Enrichment (-1*logP-value)
Cerebellum
neurotransmitter transport
exocytosis
neurotransmitter secretion
secretion by cell
potassium ion transport
0 10 20 30 40 50 60 70 80
Cortex
synaptic transmission
transmission of nerve impulse
ion transport
learning
divalent metal ion transport
0 50 100 150 200
E14.5 brain
central nervous system development
central nervous system differentiation
brain development
forebrain development
pallium development
0 30 60 90 120 150
E14.5 heart
heart development
heart morphogenesis
tube development
tissue morphogenesis
blood vessel development
0 10 20 30 40 50 60 70 80
E14.5 liver
myeloid cell differentiation
cofactor biosynthetic process
erythrocyte differentiation
erythrocte homeostasis
negative regulation of cell adhesion
0 5 10 15 20 25
E14.5 limb
embryonic morphogenesis
organ morphogenesis
embryonic limb morphogenesis
limb development
skeletal system development
0 50 100 150 200 250
Heart
anatomical structure formation
vasculature development
blood vessel development
blood vessel morphogenesis
angiogenesis
0 5 10 15 20 25 30 35
Intestine
regulation of hormone levels
striated muscle cell differentiation
lipid catabolic process
columnar epithelial cell differentiation
hormone metabolic process
0 5 10 15 20 25
Kidney
protein tetramerization
retinoic acid metabolic process
cysteine metabolic process
nephron development
epithelial cell development
0 5 10 15 20
Liver
organic acid metabolic process
carboxylic acid metabolic process
cellular ketone metabolic process
lipid metabolic process
monocarboxylic acid metabolic process
0 30 60 90 120 150
Lung
postive regulation of smooth muscle cell proliferation
vascular endothelial growth factor recepotr signaling pathway
regulation of smooth muscle cell proliferatin
cyclic nucleotide catabolic process
gastrulation
0 5 10 15 20 25
MEF
tissue morphogenesis
morphogenesis of branching epithelium
gland morphogenesis
morphogenesis of branching structure
extracellular matrix organization
0 10 20 30 40 50 60 70 80
mESC
stem cell differentiation
stem cell maintenance
stem cell development
embryonic pattern specification
trophectodermal cell differentiation
0 5 10 15 20 25
Olfactory bulb
metencephalon development
regulation of cellular component size
di-,tri-valent inorganic cation transport
divalent metal ion transport
calcium ion transport
0 5 10 15 20 25
Placenta
placenta development
embryonic plancenta development
developmental maturation
cell matruation
regulation of osteoblast differentiation
0 5 10 15 20 25 30
Spleen
cell activation
leukocyte activation
lymphocyte activation
B cell activation
regulation of immune system process
0 20 40 60 80 100
Testis
embryonic pattern specification
localization with membrane
axis specification
male genitalia development
genitalia development
0 3 6 9 12 15
Enrichment (-1*logP-value)
Thymus
T cell differentation
lymphocyte differentiation
T cell activation
leukocyte differentiation
lymphocyte activation
0 50 100 150 200

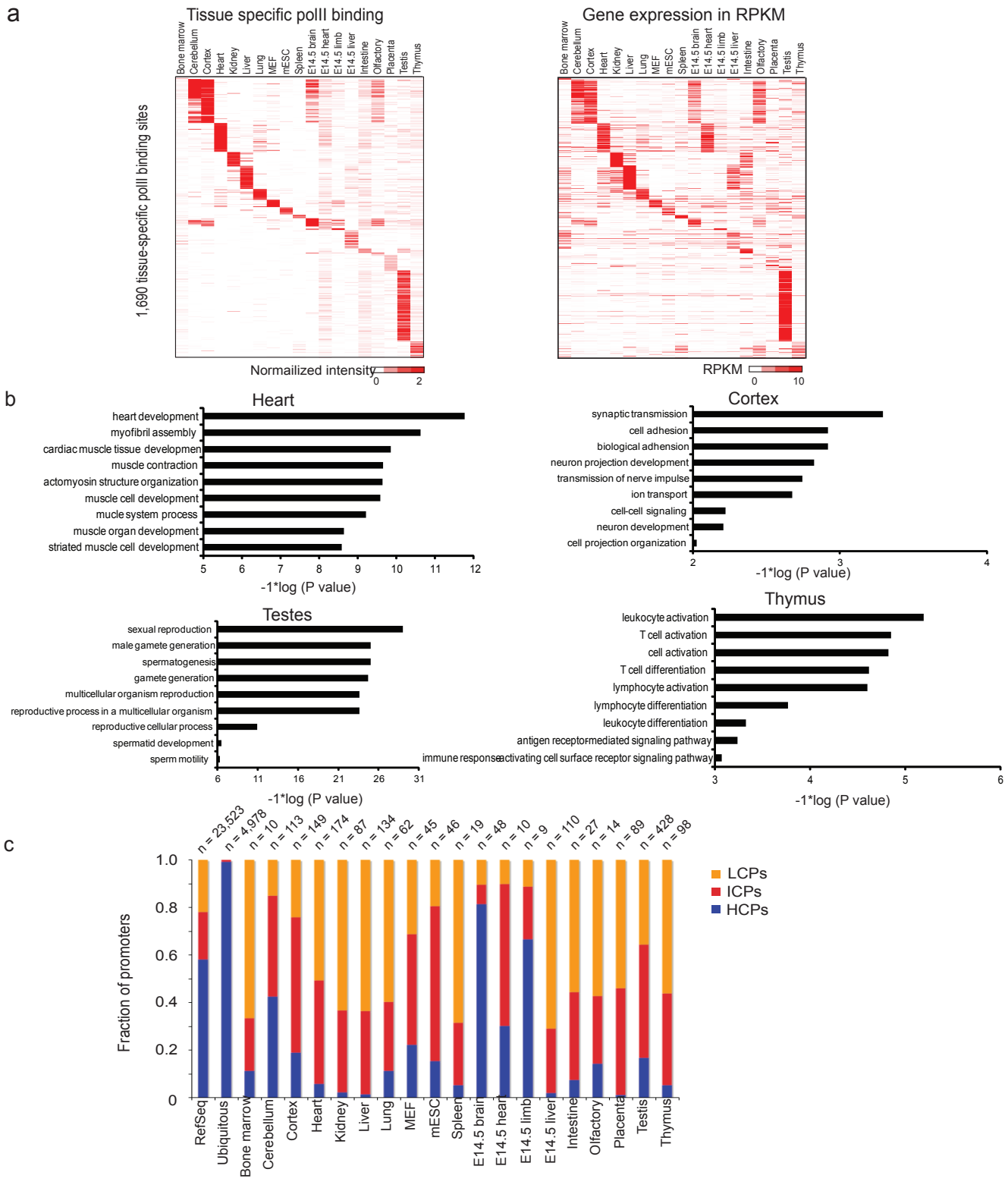
Supplementary Figure 11: Gene Ontology analysis for the 19 clusters of tissue-specific. Tissue-specific enhancers were identified based on the normalized H3K4me1 intensities (Fig. 4e). The complete enhancer list from each cluster was used as the test regions and the background regions were set as the whole genome. We used the default settings of GREAT and removed the redundant GO categories.



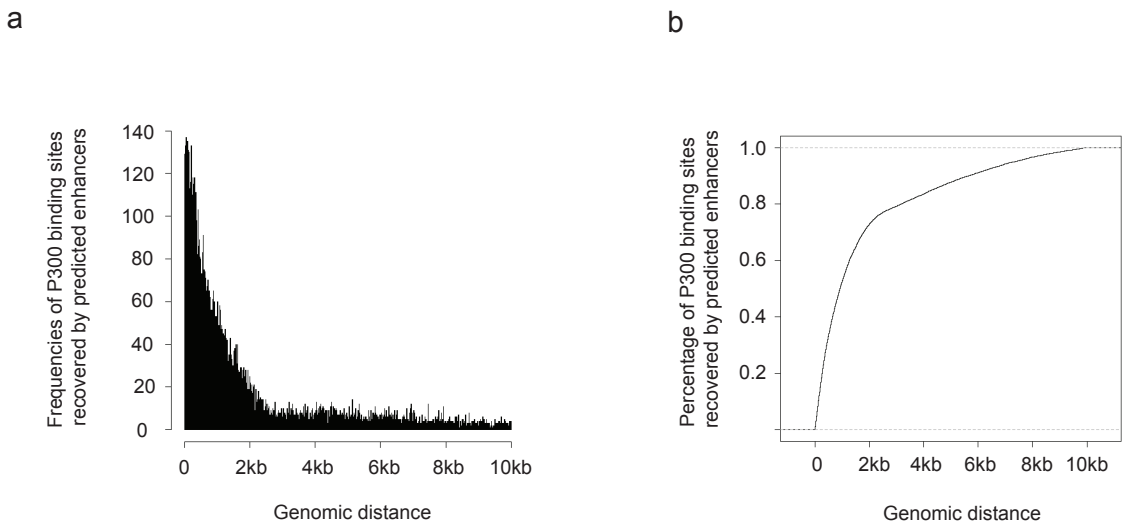
Supplementary Figure 12: Chip-Seq datasets are highly reproducible. a, Correlation between two biological replicates of H3K4me3 binding in mESCs. We divided the mouse genome into 1000 bp bins and counted the number of reads in each bin. Then we computed the Pearson correlation coefficient between the two vectors of the normalized tag counts. **b**, Distribution of Pearson correlation coefficients between biological replicates. c, Schematic process of identifying promoters, enhancers and CTCF binding sites. Please refer to methods for detailed information.



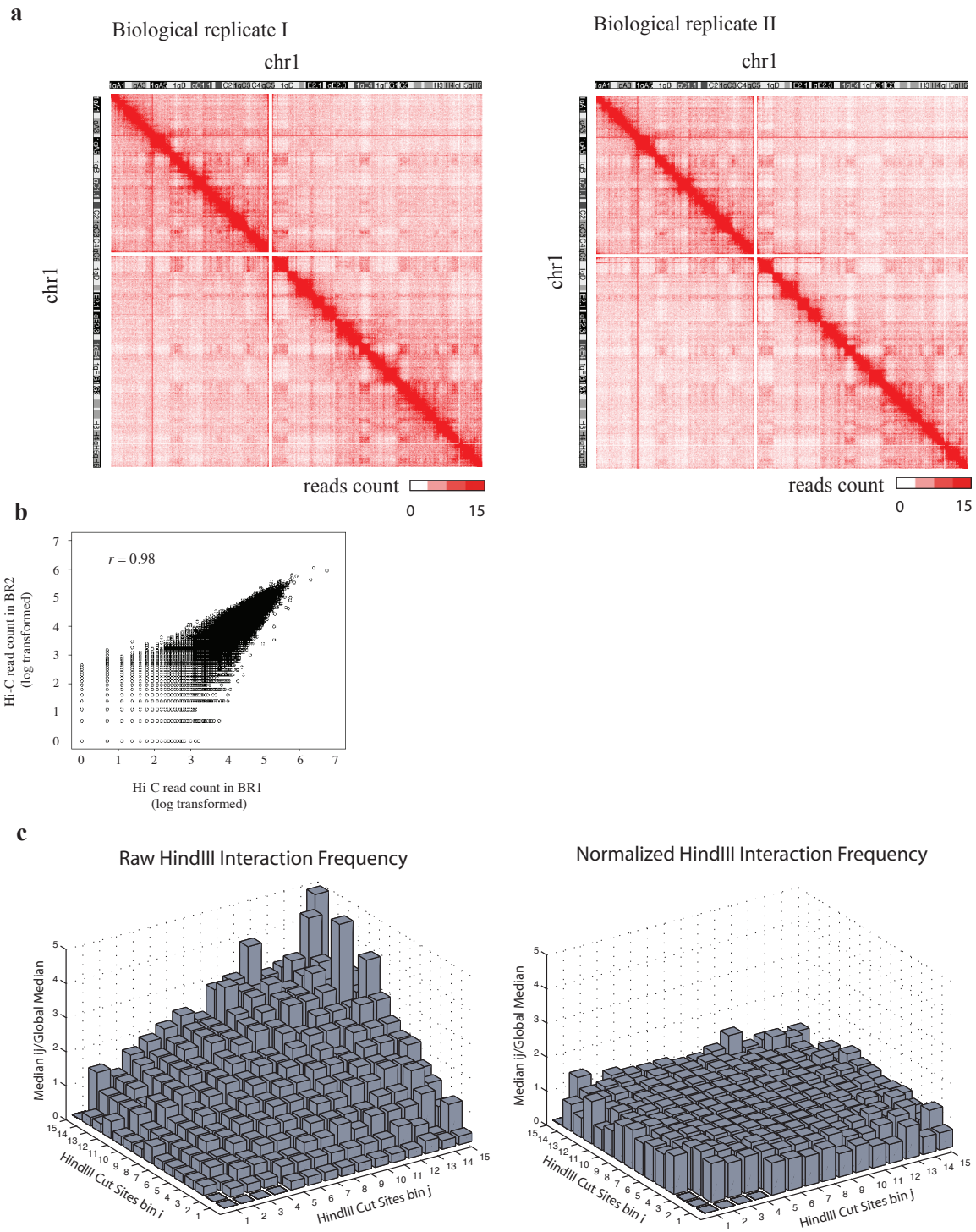
Supplementary Figure 13: Analysis of the promoter-distal p300 binding sites in the mouse genome in selected tissues. a, Chromatin state at promoter-distal p300 binding sites showing an enrichment of H3K4me1 but depletion of H3K4me3 signals. **b**, Comparison of promoter-distal p300 binding sites with enhancers predicted with chromatin state in the heart, kidney, liver, lung and mESCs. **c**, Percentage of predicted enhancers associated with H3K27ac in 19 tissue and primary cell types.



Supplementary Figure 14: Tissue-specific polII binding at RefSeq annotated promoters. a, Heatmaps showing tissue-specific polII binding sites at promoters and the corresponding gene expression patterns in 19 tissue and primary cell types. **b**, Gene ontology analysis of the genes associated with tissue-specific polII binding at promoters in heart, cortex, testis, and thymus. **c**, CpG content analysis of promoters with tissue-specific polII binding, promoters with ubiquitous polII binding, and all of the RefSeq promoters.



Supplementary Figure 15: Distribution of the distances between predicted enhancer to the nearest p300 binding sites. a, Frequencies of p300 binding sites recovered by the predicted enhancers at different genomic distances. **b**, Cumulative percentage of p300 binding sites recovered by the predicted enhancers at different genomic distances.



Supplementary Figure 16: Analysis of long-range chromatin interactions in adult cortex. a, Heatmaps showing the interaction frequency by Hi-C analysis along chromosome one. Normalized read counts in 200Kbp bins are shown. **b**, Hi-C experiments are highly reproducible. The Pearson's correlation coefficient is 0.98 between the contact matrices for two biological replicates. **c**, Bias plots for Hi-C interaction data. The two horizontal axis represent the number of HindIII cut sites in each of two interaction bins, i and j. The vertical axis is the median number of Hi-C interacting reads between all bins i and j with the given HindIII cut site frequency divided by the global median. Perfectly unbiased data should be equal to 1. The raw Hi-C data shows clear bias, as increasing HindIII cut site frequency is generally correlated with higher interaction frequency. The normalized data shows no bias.

Supplementary Table 1:Number of uniquely mapped monoclonal reads for each ChIP-Seq experiments.

		Replicate 1	Replicate 2
H3K4me1	Bone Marrow	13,373,224	20,395,300
	Cerebellum	14,671,683	10,440,493
	Cortex	16,323,654	10,070,479
	Heart	24,418,456	15,327,478
	Kidney	12,804,687	10,466,062
	Liver	13,377,730	7,324,463
	Lung	13,512,264	10,223,076
	MEF	11,668,843	10,762,982
	mESC	12,792,833	11,009,643
	Spleen	9,055,300	12,022,988
	E14.5 brain	12,432,689	23,524,417
	E14.5 heart	12,125,940	17,120,072
	E14.5 limb	11,554,987	12,093,501
	E14.5 liver	22,022,445	12,243,446
	Intestine	7,681,649	12,775,900
	Olfactory	9,911,792	12,630,337
	Placenta	9,702,241	10,407,572
	Testis	14,892,620	9,230,110
	Thymus	7,883,456	9,340,395
H3K4me3	Bone Marrow	12,124,246	12,110,310
	Cerebellum	12,805,071	8,861,655
	Cortex	10,160,758	7,740,312
	Heart	6,165,942	9,648,271
	Kidney	12,986,624	5,905,400
	Liver	6,224,743	7,767,644
	Lung	12,372,534	4,831,577
	MEF	8,268,131	9,281,640
	mESC	12,328,541	12,216,043
	Spleen	8,308,524	5,637,080
	E14.5 brain	13,289,643	12,403,002
	E14.5 heart	9,643,633	9,391,219
	E14.5 limb	18,397,120	13,184,489
	E14.5 liver	16,777,841	5,319,531
	Intestine	15,564,829	15,149,239
	Olfactory	6,336,829	10,153,515
	Placenta	5,625,681	5,568,862
	Testis	21,578,802	10,354,684
	Thymus	12,643,699	14,534,407
CTCF	Bone Marrow	12,132,382	9,021,689
	Cerebellum	10,015,871	18,039,743
	Cortex	7,672,964	7,158,291
	Heart	8,515,104	8,268,562
	Kidney	8,653,666	10,926,498
	Liver	15,050,731	
	Lung	9,171,792	9,916,575
	MEF	7,708,567	8,856,791
	mESC	10,346,108	8,335,793
	Spleen	8,873,258	7,775,015

	E14.5 brain	10,689,875	8,397,747
	E14.5 heart	2,803,774	
	E14.5 limb	8,601,956	4,429,302
	E14.5 liver	5,222,802	16,396,462
	Intestine	10,659,546	19,373,518
	Olfactory	10,666,767	17,259,107
	Placenta	5,096,034	
	Testis	5,339,734	11,082,450
	Thymus	2,012,918	7,466,036
polII	Bone Marrow	10,500,424	15,345,277
	Cerebellum	5,336,278	7,990,676
	Cortex	10,195,208	7,357,268
	Heart	7,082,141	8,963,213
	Kidney	12,841,747	9,155,910
	Liver	10,987,805	11,900,992
	Lung	7,793,148	11,876,951
	MEF	11,163,405	16,712,650
	mESC	9,649,642	6,292,363
	Spleen	16,043,631	14,505,404
	E14.5 brain	12,135,505	9,575,297
	E14.5 heart	3,880,881	
	E14.5 limb	6,068,957	6,679,821
	E14.5 liver	8,273,907	
	Intestine	13,111,225	30,091,514
	Olfactory	6,287,618	9,222,151
	Placenta	4,033,436	
	Testis	13,111,225	12,790,900
	Thymus	3,751,689	6,696,985
Input	Bone Marrow	8,636,208	11,240,181
	Cerebellum	9,097,822	12,507,140
	Cortex	14,578,545	8,405,851
	Heart	7,291,052	9,570,308
	Kidney	13,457,266	8,634,862
	Liver	12,732,885	10,566,370
	Lung	9,977,762	3,480,103
	MEF	10,050,931	13,143,875
	mESC	10,229,779	12,809,138
	Spleen	11,672,482	18,981,865
	E14.5 brain	21,225,906	9,282,280
	E14.5 heart	7,291,052	6,976,597
	E14.5 limb	12,510,880	7,220,643
	E14.5 liver	11,102,055	12,676,498
	Intestine	17,659,642	18,463,860
	Olfactory	7,789,689	11,774,691
	Placenta	3,728,117	8,590,542
	Testis	3,117,138	4,767,325
	Thymus	7,777,879	8,500,794
H3K27ac	Bone Marrow	8,169,499	10,415,298
	Cerebellum	9,104,283	8,476,234
	Cortex	6,966,005	7,285,530
	Heart	10,231,666	9,674,459
	Kidney	9,251,927	9,722,804

	Liver	12,005,280	8,868,817
	Lung	8,740,346	
	MEF	8,502,723	6,648,587
	mESC	7,326,385	7,225,895
	Spleen	7,844,644	9,479,069
	E14.5 brain	10,862,944	7,017,081
	E14.5 heart	9,757,603	9,199,816
	E14.5 limb	9,358,468	7,059,078
	E14.5 liver	9,186,844	10,103,390
	Intestine	7,097,544	10,021,883
	Olfactory	7,302,031	9,703,539
	Placenta	13,392,003	10,411,954
	Testis	9,291,645	6,064,761
	Thymus	7,092,986	21,962,027
P300	Heart	8,780,662	11,323,531
	mESC	10,013,496	10,834,923
	liver	11,517,882	
	lung	13,143,468	
	kidney	7,278,957	

Supplementary Table 2: Pearson correlation coefficient for each pair of biological replicates in ChIP-Seq and RNA-Seq experiments.

	CTCF	H3K4me3	polII	H3K4me1	H3K27Ac	RNA-Seq
BoneMarrow	0.93	0.97	0.78	0.81	0.88	0.98
Cerebellum	0.83	0.99	0.85	0.85	0.93	0.97
Cortex	0.94	0.98	0.90	0.90	0.94	0.98
E14.5 brain	0.87	0.99	0.79	0.86	0.97	0.96
E14.5 heart		0.98		0.89	0.96	0.94
E14.5 limb	0.75	0.99	0.78	0.86	0.96	0.99
E14.5 liver		0.98		0.95	0.98	0.97
Heart	0.89	0.97	0.93	0.93	0.97	0.91
Intestine	0.70	0.99	0.91	0.76	0.95	0.90
Kidney	0.93	0.98	0.76	0.88	0.96	0.99
Liver	0.92	0.99	0.82	0.92	0.98	0.96
Lung	0.90	0.98	0.88	0.91		0.99
MEF	0.92	0.99	0.94	0.90	0.95	0.94
mESCs	0.94	0.99	0.82	0.93	0.97	0.99
Olfactory Bulb	0.85	0.99	0.84	0.94	0.94	0.99
Placenta		0.99		0.92	0.97	0.99
Spleen	0.87	0.94	0.93	0.69	0.95	0.97
Testis	0.68	0.83	0.68	0.90	0.96	0.99
Thymus	0.88	0.99	0.91	0.97	0.93	0.99

Supplementary Table 3: Numbers of expressed genes by BioGPS in each tissue whose promoters are not recovered in this study.

Tissue	Number of transcribed genes whose promoter were not recovered
uterus	1819
epidermis	352
stomach	331
testis	305
adipose_white	254
ovary	253
umbilical_cord	252
intestine_small	252
placenta	248
retinal_pigment_epithelium	248
prostate	247
mammary_gland__lact	233
cornea	228
mammary_gland_non-lactating	227
salivary_gland	225
retina	224
ciliary_bodies	222
adipose_brown	220
skeletal_muscle	219
dorsal_striatum	218
bladder	217
lacrimal_gland	215
bone	213
kidney	213
eyecup	212
iris	209
cerebral_cortex_prefrontal	207
lens	203
spleen	203
intestine_large	200
osteoblast_day14	199
mast_cells	198
spinal_cord	197
lung	196
dorsal_root_ganglia	196
pancreas	193
mast_cells_IgE+antigen_1hr	193
mast_cells_IgE+antigen_6hr	193
osteoblast_day21	191
lymph_nodes	190
mast_cells_IgE	190
3T3-L1	190
pituitary	189
hypothalamus	183
heart	182

dendritic_cells_lymphoid_CD8a+	181
nucleus_accumbens	181
osteoblast_day5	180
macrophage_bone_marrow_2hr_LPS	180
hippocampus	180
cerebellum	180
MEF	180
macrophage_peri_LPS_thio_1hrs	179
adrenal_gland	179
amygdala	178
olfactory_bulb	177
min6	176
macrophage_peri_LPS_thio_0hrs	176
bone_marrow	175
cerebral_cortex	173
NK_cells	172
dendritic_cells_myeloid_CD8a-	171
T-cells_foxP3+	170
macrophage_bone_marrow_0hr	170
macrophage_bone_marrow_24h_LPS	168
neuro2a	167
T-cells_CD4+	167
liver	167
C3H_10T1_2	167
common_myeloid_progenitor	166
mIMCD-3	164
granulo_mono_progenitor	164
osteoclasts	164
dendritic_plasmacytoid_B220+	163
thymocyte_SP_CD4+	162
granulocytes_mac1+gr1+	161
macrophage_bone_marrow_6hr_LPS	161
B-cells_marginal_zone	161
microglia	161
nih_3T3	161
T-cells_CD8+	159
C2C12	159
embryonic_stem_line_Bruce4_p13	158
embryonic_stem_line_V26_2_p16	156
RAW_264_7	156
follicular_B-cells	156
stem_cells__HSC	154
B-cells_GL7_positive_Alum	153
B-cells_GL7negative_Alum	152
Baf3	150
thymocyte_SP_CD8+	150
B-cells_GL7_positive_KLH	149
thymocyte_DP_CD4+CD8+	146
macrophage_peri_LPS_thio_7hrs	145
B-cells_GL7_negative_KLH	142
mega_erythrocyte_progenitor	135

Supplementary Table 6: Table describing the general features of EPUs identified.

	0%	25%	50%	Average	75%	100%
Num. of promoters	1	1	1	2.35	3	129
Num. of enhancers	1	5	10	14.54	19	180
Enhancer/promoter ratio	0.01	2.5	5.67	9	11	139
EPU size	3,021	34,490	73,200	143,200	148,400	9,948,000

Supplementary Table 8: List of enhancer promoter pairs tested by 3C and their correlation scores.

Primer Locations		Correlation score		Predicted to be linked		Validated by 3C	
Enhancer (anchoring point)	Promoter	Promoter (polII) vs. enhancer (H4K4me1)	Promoter (H3K27Ac) vs. enhancer (H3K27Ac)	CTX	mESCs	CTX	mESCs
chr10:69,996,146-69,996,165	chr10:69,901,922-69,901,945	0.4	0.5	Yes	No	Yes	No
	chr10:69,977,279-69,977,298	0.25	0.34	Yes	No	Yes	No
	chr10:70,058,666-70,058,692	0.08	0.04	No	No	No	No
chr3:82,055,789-82,055,808	chr3:81,885,427-81,885,446	0.45	0.6	Yes	No	Yes	No
	chr3:81,950,584-81,950,603	0.41	0.55	Yes	No	Yes	No
chr12:71,375,859-71,375,878	chr12:71,329,105-71,329,124	0.17	-0.12	No	No	No	No
	chr12:71,450,905-71,450,924	0.26	0.14	Yes	No	Yes	No

SupSupplementary Table 11: List of the de novo motifs that can be matched to a known TF that has been reported to function in the same tissue

Tissue	Transcription factors	Tissue	Transcription factors
Bone marrow	PU.1, Cebpa	MEF	Jundm2, Ap-1
mESCs	Oct1, Sf1, Sox2, Tcf12, HEB	Lung	Foxk1, Foxf1a, E2f1
Cortex	Pou6f1, Rfx7, Oct1, Bach1, Cebpg, Smad3, Mef2c	Liver	Hnf6, Hnf4a, Pbx1, Cebpa, Foxa1, Pparg, Pcbp1, Foxa2
E14.5 brain	Vsx2, Pou6f1, Oct1, Lmx1a, Pou3f2, Tgif1, Sox10, Tcfcp2, Hmx2, Pax6	Cerebellum	Tcf3, Nf1, Zic1, Nr4a2, Pou3f2, Zic3
E14.5 heart	Gata5, Usf1, Tead1	Olfactory bulb	Hlx, Ap1, Prrx2
E14.5 limb	Hoxc13, Hoxa9, Gfi1b, Zfp238	Placenta	Tcfap2c, Nr2f2, Nfe2
E14.5 liver	Gata1, Gata3	Spleen	PU.1, Oct1, Irf1
Intestine	Gata6, Cdx1, Hnf4a, Cdx2	Testis	Ets1, Mybl1, Rfx2, Ahr
Kidney	Hnf1a, Hnf4a	Thymus	Zeb1, Runx1, PU.1, Ets1

Supplementary Table 12: List of enriched known motifs from Homer in tissue-specific promoters and enhancers.

Tissue	Enriched motifs found in promoter regions	Enriched motifs found in the enhancer regions
Bone marrow	Sp1, ETS,ELF1,NRF1	PU.1, CEBP, ETS1, RUNX-AML, CEBP
Cerebellum	NRF1, Sp1, JunD,	Atoh1 (bHLH), NF1, HEB, Tlx, Pdx1
Cortex	GFY, RFX, CRE(bZIP), X-box, Sp1,	RFX, X-box, Mef2a, Atoh1, AP-1
Heart	Sp1, Mef2a	ETS, EWS, ERG, Mef2a, NF1
Kidney	Sp1, GFY-Staf	Hnf1, HNF4a, RXR, PPARE, Pax5
Liver	SP1, HNF4a, CEBP(bZIP)	HNF4a, CEBP, Foxa2, TR4, PPARE, RXR, FOXA1
Lung	Sp1	ERG, ETV1, Foxa2, GABPA, FOXA1,
MEF	Sp1, ETS	Jun-AP1, Ap-1, NF-E2, TEAD, c-Jun-CRE
mESCs	Sp1, Klf4, NRF1, E2F, ELF1,ETS,	Oct4, Sox2, Nanog, Klf4, EKLF, Esrrb
Spleen	ELF1, ETS, GAPBA, PU.1, Sp1,	PU.1, ETS, GABPA, ETV1, ERG, RUNX
E14.5 brain	Sp1, Sox2,	Lhx3, RFX, X-box, Sox2, Tcf12
E14.5 heart	Sp1	Gata2, Gata1, Mef2a, TEAD, NF1
E14.5 limb	Sp1	Hoxc9, Cdx2, Atoh1, Myod, Lhx3
E14.5 liver	GFY, ELF1, ETS, GABPA, E2F, Sp1,	Gata1, Gata2, NF-E2, EKLF
Intestine	Sp1, ETS	HNF4a, Gata2, Gata1, RXR, PPARE
Olfactory bulb	NRF1, RFX, REST-NRSF	Jun-AP1, AP-1, Lhx3,
Placenta	ELF1, ETS, Sp1, YY1,	AP2gamma, AP-2alpha, TR4, RXR, AP-1
Testis	RFX, ELF1, Sp1, X-box, Rfx1, NRF1, JunD,	NRF1, ELF1, CTCF, RFX, USF1

Supplementary Table 13: Primers sequences and chromosome locations of MEF-specific, mESC-specific enhancers and random genomic regions used for enhancer reporter assay.

ID	Forward primer	Reverse primer	Genomic location
M1	GCCACACACCCAGTACCTCT	CGGCTGAGGTCTCTTCTGAC	chr18:60,829,430-60,831,096
M2	GTCCAACAAGAGGGGATTCA	CACCCTAGCCTTCAGCAAAC	chr6:88,875,251-88,876,750
M3	TAAAATCACAGAAAAGCCCAGA	CCCGTGACCTAGTGTTTCTGA	chr18:60,823,240-60,824,215
M4	TGCCTCAGTTTCCTGGTTTC	GGCGCATGTGAACATACAGA	chr2:27,717,123-27,718,633
M5	CCTGTCCTCTATCGCCTCAG	GCCTGTCAACTGTGCAGAAA	chr8:13,530,275-13,531,941
M6	AAGGCTGATTGCCTCCTTCT	TAGCATCCTCGCCAGTCTTT	chr8:124,709,959-124,712,180
M7	CACACTCAGGGTAGCAGCAA	TCCTGTCGCTGCATTCATAG	chr15:85,516,233-85,518,454
M8	ACCTCCAAGCTCAGCAGCTA	CGGAAGGTTTCCTGTCATGT	chr14:48,908,280-48,909,812
E1	TTGGGTCATGGCTTCTTAGG	GTGCAAGGCTGGAGACTCAT	chr15:51,843,951-51,846,000
E2	GCTGGAGGAAAAGACAGTGC	GAGGGTCCACCATACACACC	chr2:30,335,445-30,337,666
E3	AGCCAACATCCGCTCATAAC	AGCTAAGCCCCAGTCCTCTC	chr8:122,880,274-122,882,273
E4	GGAGGTCAACGTCTCTGCTC	AGCACCAGGGTTGTTGTTTC	chr5:129,752,508-129,754,576
E5	GGGCTCACTAGCCTGCAATA	CTTCATGCTTGCTCCTCTCC	chr8:73,160,067-73,162,088
E6	TCTAGCATCCATCCCTGTCC	AGTGTGGCCATTGGTAGGAG	chr15:84,487,586-84,489,207
R1	CTTCAGTGGGAGAGTCAGTGG	TGCAGGTGTGTGGGTAAGC	chr18:60,476,744-60,478,744
R2	CAACAGCTTTGAAACCCTGA	TCATGCCTCCTGTGGTGATA	Chr1: 7940500- 7942500
R3	GGACTTCCAGGTTCCTCACA	GCCATTTTTCAATGCAGGAG	chr1:27,739,422-27,741,755
R4	TGCCTCACAATGGAACTGAA	AGGCCACTTTCTGAAAAGCA	chr3:10,996,456-10,998,013
R5	AAGGACCCAGCCTGTGAGTA	ACCATGATTCTCTGGCTGCT	chr5:6,955,071-6,957,539
R6	TTTCCACTGGGGCTGTTAAA	ACAGAAGCAAGGCCACAGAT	chr6:10,434,517-10,436,532

Supplementary Table 14: Primers sequences and chromosome locations of novel promoters predicted MEF, mESCs, and random genomic regions used for promoter reporter assay.

ID	Forward primer	Reverse primer	Genomic location
MEF1	CCCCATGGCTACTAGCAAGA	TCAACAAGACCCGGGTAACT	chr1:72,282,608-72,283,862
MEF2	GGCTCGCGCATATTAACACT	CCCTGCTAATTGGCTCTCTG	chr11:120,176,972-120,178,515
MEF3	CCACTGAGCCATCTTTCTCC	ACGACGGCTTTTGTTGTACC	chr1:88,383,424-88,384,740
MEF4	TAGCACATTCCCCTGTCCTC	CCCTGAAGCACTCTGCTACC	chr2:32,821,808-32,823,140
MEF5	GCATTTGGAAAGGATTTGGA	TGAAACACCCCCACGTTATT	chr8:46,019,901-46,021,052
MEF6	AGGTTTTGGATGCTTGTTCG	GGCGGTGTCTGGAGAGTAGA	chr4:133,466,066-133,467,413
MEF7	TTCTGCTTAAGTTCCTGAAGTTTTT	TGCGCGGAGTTAACTGTAGA	chr10:81,640,535-81,641,906
MEF8	AGCCTTCACCTTTGCACTGT	AGCTGTCTGCAACATCATGG	chr12:113,931,008-113,932,428
MEF9	GCTGCTGAACAGAACCTTCC	AAGAACCTGTTCCGCACTGT	chr16:21,332,693-21,334,078
MEF10	GTACCGTTCGGTCCCTACAC	GGGTGCTTTGAGATTTTCGT	chr6:117,829,050-117,830,421
MEF11	CAAGTCAAAGCACACACAGGA	CAACAGCTCTGTGCATGTGA	chr7:88,667,757-88,668,832
MEF12	CAAGTGCCAGACCAGTTTGA	TTGATCCCATTTTCCCAGAG	chr1:4,561,098-4,562,521
MEF13	GGTGATGCTTTCCTGGGTTA	CAGATCCCGCCTCTCTACTG	chr13:23,487,516-23,488,786
MEF14	CCAGGGTAGTAAAATGTCTTCTGTG	AGATGGTGCCCTTTTTGTTG	chr1:172,963,082-172,964,359
MEF15	ACCCAAACACGACACCATCT	CTCCTTGCACACCCTGTTTT	chr19:61,275,284-61,276,321
MEF16	GCTGCTCGTTGGAGTAGACC	CCACAAGACAAAATGCTCCA	chr13:21,366,741-21,367,942
MEF17	GTAGCTTCGCTCCCTGACAC	ATCGGGTTTAGCAGAGCAGA	chr11:37,049,105-37,050,498
MEF18	AGAGGCAGGTGCTCAGAAAA	CTGCCTGGTTGTGGAGATTT	chr14:51,702,299-51,703,553
MEF19	GGACTTCAGATTCCCCCAAG	GTGGTGTCAGGTGCTGTGAC	chr4:88,513,577-88,514,670
MEF20	GCTACTGCTGCTTCCAAACC	CACTTAGTGGGGAGGAGAGG	chr8:52,000,477-52,001,938
MEF21	GGAGGGAGGATGAAGTAGGG	AGCCTGGAGGAAGCTTTAGG	chr6:70,962,454-70,963,748
MEF22	ACACACGACAACCAGCAAAG	ACCTAGCCCTGTGTGTCGAG	chr6:31,037,357-31,038,733
MEF23	TCCATCCACATAAGGGTGAG	GGCATTTTCCAAGCTGAATG	chr10:5,056,630-5,057,713
MEF24	ATCTCCGGAAGCCCTAACTC	GGAAGGAAGGCAAAGGAAAC	chr6:4,406,665-4,407,901
MEF25	TTTCCAGTTGGTGGATGACA	AAGCAGCAACAGCACATCAC	chr2:30,319,171-30,320,263
MEF26	TCAACTCCCAGCACTTAGCC	AGGCTTAGTCCAGTCCACCA	chr1:184,449,607-184,450,750
MEF27	GCTTCAGTCTCCATGTTCCTG	AGCAAAAGCCAGAATCTCCA	chr11:94,793,896-94,795,214
MEF28	AGTTCTCAATGCTGGGCAAC	CCCCCAGCAACAGTCAAT	chr1:173,192,458-173,193,486
MEF29	TCCCCCAATTTTTCTCTGTG	GCATGGAATTACGCTGTGTG	chr6:4,438,664-4,440,057
MEF30	TCCATTTGCAGTCAGTGGAG	TCCAGTTCTGCGTCTTCCTT	chr7:106,503,912-106,505,303
MEF31	CGTTAGAGCCAGAAGCCAGT	CGCCCTACACCATAACCAAT	chr12:70,602,990-70,604,165
MEF32	GGCAAGGCAAACACTACCAT	CCCTTTTCTAGCCTGCCTTC	chr8:131,177,663-131,179,109
MEF33	CAGTCCTGAACAGCGACAAT	CAGGCGGTCTCCTAAAAATG	chr6:31,037,647-31,038,728
MEF34	CAGGGTCAGTGAGCTTGACA	AATCCCACGTACAGGCTTTG	chr14:118,400,191-118,401,401
MEF35	AGTCGAAAAACTGCCACCAT	GGGAACGACAACAACAACAA	chr5:13,628,493-13,629,571
MEF36	ACCCCACCAAGGAACATACA	GATTCTAGCGGGGTCTAGGG	chr1:156,730,738-156,732,002
MEF37	CTACCCCAGCTTCCACAAAA	CGATCAGACTGGGATTTGCT	chr14:68,230,847-68,232,082
MEF38	CTGGTCAGCACGCATAACAT	ACCTGTCTCTCCCCAGTGTG	chr10:126,246,319-126,247,564
MEF39	TTTTTGCATGGAAAGCCAGT	CCACATGAAAAACAGAGTTTGC	chr2:65,656,036-65,657,085
MEF40	ATCTCGGGTTCTGGTGACTG	GGTTTATGGCGTGCTGACTT	chr14:87,967,514-87,968,917
mESC1	CATTCACTTTGGTGGGCTCT	TCATTGGGCTAATGTCAAAGG	chr2:75,471,401-75,472,561
mESC2	AGTCGAAGGTCATGGGTTTG	AACACCACCGCTCACCTC	chr12:18,393,657-18,394,810
mESC3	CTGAAACCCACACTCCCATC	TGGAACTGAAGGAACCCAAG	chr15:100,922,977-100,924,412
mESC4	CACTGCCGGAAGGTTAGAAG	GCTGGCCTTAGGAGTTCAGA	chr2:166,903,330-166,904,794
mESC5	CTCAGGCGGTCCTAAGAATG	TAGCACTGTGCGTTTGCTCT	chr4:138,262,409-138,263,533
mESC6	AGCTCAGACCACACCGTTCT	GCTATGCCTGGCTATCTTAGTTC	chr17:30,118,145-30,119,424
mESC7	AATCTGACCGCCAATAGCTG	GTTGACTCTGGCAGGGACTG	chr8:109,632,424-109,633,774
mESC8	GCTGGGACTCTGAGAACTGG	TGAGTGCAGAGAGGTCATGG	chr4:154,028,643-154,030,048
mESC9	CGTCTGCATCTGTTTTGTGG	GGACACTGATCCGTCCAGTC	chr8:73,214,953-73,216,209
mESC10	TGGGCAGGACTTATTCAACC	AGAGGGCCACAGCCTAAAAT	chr9:114,572,308-114,573,743
mESC11	CTGCAAGTTGAATCCTCAGC	AAAAGTTGGGATGGGAGGTC	chr7:50,508,728-50,509,835
mESC12	CCCTGGTTGGCACATTACTT	CTGGGCACCCTTCCTCTTAT	chr5:110,514,708-110,515,903
mESC13	GCTACAGCCATAGAATCCAATTTT	AAATGCTCGGAGCTGAAAGA	chr8:91,576,369-91,577,522
mESC14	AACAGGAGCAGCATGGAGAT	ACAATGGAGCAGAGGTGTCC	chr14:76,915,379-76,916,761
mESC15	TGGGCAATAAGAGCTGGACT	TGGTTGGTTGGGTTTTGG	chr4:133,512,988-133,514,350
mESC16	CCCACGGTATGGAATAATCG	CAGACCCTGCCATACTGGAC	chr9:113,954,352-113,955,827
mESC17	CATCTCCTCTGCCTTTGACA	TCTCATTCCGCTTTTAAACCA	chr2:167,088,587-167,089,586
mESC18	TATGGACCGAAGCACAAACA	AATTGTTCCTGATGGCGAAG	chr19:55,803,480-55,804,780
mESC19	CAGGGACAGGGTTAAGAGCA	ACTGCGGTTTGAGGTGAGTT	chr5:33,876,910-33,878,336
mESC20	AGGAGGCGCCTAACACTTCT	CGTCATTCCTCAAACCTGCT	chr8:28,332,972-28,334,480
mESC21	GCAGGATCTGACTTGGGGTA	GGGCTATTGGGAGGTTTAGG	chr4:140,840,789-140,842,238
mESC22	TGAACTTGGGAGGGGTACAG	GAGGCATTGAAAGCATCTGG	chr4:44,965,616-44,967,013

mESC23	CTCCAAGTGCCTCAGTAGCC	TCCCTCCAGACTTTCCACAC	chr17:37,107,047-37,108,640
mESC24	GAGCCCTCACTCCAGTCCTA	ATATCTAGGCGGCCGTGTC	chr3:96,373,157-96,374,474
mESC25	GGCTGCATATAACTCAACACCTC	GGTTCCAGCCACTCAGGTTA	chr17:29,655,802-29,657,008
negative1	AGCAGACCCTGTTTGACCAC	TGTCTGATTTCCCAGGCTAAA	chr1:4,026,193-4,027,420
negative2	GCCACATCTTTTGCATCTCA	GAAGCTCAAGCAAGCTCTCC	chr2:4,004,617-4,005,736
negative3	TTTGTGCCCAAACTGAAATATG	GAATGTAGTGGGTGTGCAAATG	chr3:6,326,256-6,327,494
negative4	CCATCAAGAAACAGCAGCAA	TTTCTTGCCTTCTAGCTCAGG	chr4:7,455,799-7,456,990
negative5	GCACTGTGCAGAAGAGGTCA	TTGAGAAAGGCACAGGACTTC	chr6:7,362,234-7,363,436
negative6	GCCATATCCAATACTTGCAGAA	TCATGCCTCCTGTGGTGATA	chr1:7,940,959-7,942,267
negative7	GTTTGCCAAGGAACTCTTGC	TTTTGTGGTCTGTTCCAGCA	chr3:10,996,564-10,997,994
negative8	GACCTATGAACTGGATCATTGAAA	ACAGAAGCAAGGCCACAGAT	chr6:10,434,672-10,436,122

Supplementary Table 15: List of 3C primers and their location based on mm9

Primer Name	Restriction Enzyme	Chromosome	Sequence
Gucy1a3 1* F	HindIII	chr3:82,055,789-82,055,808	TCCATTCCAGATTCCCAAAA
Gucy1a3 3 F (P1)	HindIII	chr3:81,885,427-81,885,446	AAGCAGGGCTTGTGAAATGT
Gucy1a3 4 F	HindIII	chr3:81,867,162-81,867,183	CCATGTTCTGTTCTTCGAAATG
Gucy1a3 5 F	HindIII	chr3:81,877,229-81,877,248	GGACAAGGTTCTGGCTTCAA
Gucy1a3 6 F	HindIII	chr3:81,888,661-81,888,683	ATCTGAGGCTAAAGCATTCTAGG
Gucy1a3 7 F	HindIII	chr3:81,897,190-81,897,209	GAGCTCACTTGGGAGGGAGT
Gucy1a3 8 F (P2)	HindIII	chr3:81,950,584-81,950,603	GCTGGTTCTGTTTTGGGTTT
Gucy1a3 9 F	HindIII	chr3:81,942,233-81,942,252	TCTGCTAAGAAGGGCATCGT
Gucy1a3 10 F	HindIII	chr3:81,947,035-81,947,054	CGACGCACCACACACTTCTA
Gucy1a3 11 F	HindIII	chr3:81,957,261-81,957,280	TCTCAGAAAATGCCCCATGT
Gucy1a3 12 F	HindIII	chr3:81,963,999-81,964,018	CTGGGTTTTGTCATCACTGC
Gucy1a3 13 F (P3)	HindIII	chr3:82,165,307-82,165,330	CCAAGTTGTTGTCTAGAAGCAGAA
Gucy1a3 14 F	HindIII	chr3:82,150,513-82,150,534	GGTGGCCAGAATAGTTTAGAGG
Gucy1a3 15 F	HindIII	chr3:82,156,330-82,156,349	CGAATCTCTGTCCCTCCTCA
Gucy1a3 16 F	HindIII	chr3:82,166,323-82,166,342	ACCATGGCCAAAATGACCTA
Gucy1a3 17 F	HindIII	chr3:82,170,313-82,170,340	TGAGATATGTATTAACTTGCAAAAATTG
Trim9 1 F* (E1)	HindIII	chr12:71,375,859-71,375,878	TATGGATTGGCCACGGATAC
Trim9 2 F (P1)	HindIII	chr12:71,450,905-71,450,924	CAGTTTGAAAATGCCGGATG
Trim9 3 F	HindIII	chr12:71,444,104-71,444,123	TAGCAGCACAAACACGGAAG
Trim9 4 F	HindIII	chr12:71,435,902-71,435,921	ACAGGACAATGGGGAGTACG
Trim9 5 F	HindIII	chr12:71,453,802-71,453,821	GAGTGTCTTCTGCCTGATGC
Trim9 6 F	HindIII	chr12:71,454,965-71,454,985	TCATAGGTACCGGACCATAGC
Trim9 7 F (P2)	HindIII	chr12:71,329,105-71,329,124	TGGCCACAGTTGGTGTAAAA
Trim9 8 F	HindIII	chr12:71,327,939-71,327,961	CTTCCCTCTCCTTTCCTTAAACA
Trim9 9 F	HindIII	chr12:71,316,981-71,317,001	CTCAGGAGACCGCAGTTCTAA
Trim9 10 F	HindIII	chr12:71,336,400-71,336,419	TATGGGGACACCTTCTGGAG
Trim9 11 F	HindIII	chr12:71,342,315-71,342,339	ACGGTAAGAATAGCTACTGATGCTC
Fam13c 2* R	HindIII	chr10:69,996,146-69,996,165	TCCTGCTGGCAGCCTAAATA
Fam13c 3 R (P1)	HindIII	chr10:69,901,922-69,901,945	GGAGGAAAAGACTAGTTCTCCACA
Fam13c 4 R	HindIII	chr10:69,893,016-69,893,035	AGGAGGAAGGAGGGGAGAAA
Fam13c 5 R	HindIII	chr10:69,890,512-69,890,531	TCGGTGTCCTGACATCACTG
Fam13c 6 R	HindIII	chr10:69,904,796-69,904,819	AAGTAGTGGGATACACAACTTTGC
Fam13c 7 R	HindIII	chr10:69,908,707-69,908,726	AGACCAAGAGGCTTCCTGAC
Fam13c 8 R (P2)	HindIII	chr10:70,058,666-70,058,692	TCTGACTCTTGTCATGTTTTTATTACA
Fam13c 9 R	HindIII	chr10:70,052,678-70,052,703	TGTAATACTGCTTTATGAAAGTCACA
Fam13c 10 R	HindIII	chr10:70,053,244-70,053,263	TACACTGGGTGGGAAGGAAG
Fam13c 11 R	HindIII	chr10:70,071,488-70,071,510	AGTCAACATGTCTGTTTTTAGGC
Fam13c 12 R	HindIII	chr10:70,068,297-70,068,317	CCTGCATTTGCAAAAGAAACA
Fam13c 13 R (P3)	HindIII	chr10:69,977,279-69,977,298	CTGAAACCATGAGCCAGTCA
Fam13c 14 R	HindIII	chr10:69,970,908-69,970,925	GCTGCTCTGGCAAAGGAC
Fam13c 15 R	HindIII	chr10:69,966,309-69,966,330	AAAAGCATATCCCCTTTGAACA
Fam13c 16 R	HindIII	chr10:69,987,866-69,987,885	TATTTCATGACACCCCAGCA
Fam13c 17 R	HindIII	chr10:69,992,237-69,992,256	ACCTGGTGGATTCTGCTGAG

* Anchoring Primer

Supplementary Table 16: The distribution of 373,169,847 uniquely mapped paired-end reads combined from the two Hi-C experiments.

The ligation efficiency was calculated based on the number of interactions that are either >20kb for intra-chromosome reads or inter-chromosome reads.

	Total Reads	Intra-chromosome reads (dist<20kb)	Intra-chromosome reads (dist>20kb)	Inter-chromosome reads	Ligation percentage
BR1	211,738,157	109,167,702	28,391,596	74,178,859	48.44%
BR2	161,431,690	63,618,156	31,404,190	66,409,344	60.59%
Combined	373,169,847	172,785,858	59,795,786	140,588,203	53.70%